

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 4, 2003, 02:34:22 ; Search time 1672 Seconds
(without alignments)
3550.821 Million cell updates/sec

Title: US-10-069-544-2
Perfect score: 1043
Sequence: 1 MKIRVKWPMWAMGLVLSAC.....EIKSYDKLSDYKLSDYK 204

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh
-Q=/cg21/USFTO.spool/US10069544/runac_30062003_091537_26261/app_query.fasta_1.391
-DB=GenEmbl -GPM=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US10069544.@CGN 1.1.1687 @runac_30062003_091537_26261 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_ncg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pac.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pih.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hcg_hum.*
40: em_hcg_mus.*
41: em_hcg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1043	100.0	612	6 AX081151	AX081151 Sequence
2	1043	100.0	615	6 AX081149	AX081149 Sequence
3	1043	100.0	33140	6 AX067448	AX067448 Sequence
4	151.5	14.5	5628	1 AF074396	AF074396 Desulfo
5	136.5	13.1	11161	1 AE013738	AE013738 Yersinia
6	136.5	13.1	204050	1 AJ414154	AJ414154 Yersinia
7	136.5	13.1	301550	1 AP003192	AP003192 Clostridi
8	132.5	12.7	10065	1 AE000799	AE000799 Mechanoba
9	124.5	11.9	4097	3 CEU77412	CEU77412 Caenorhabdi
10	123	11.8	11268	1 AE006237	AE006237 Pasteurel
11	122	11.7	11896	1 AE000762	AE000762 Aquifex a
12	121.5	11.6	15059	1 AE000777	AE000777 Aquifex a
13	120.5	11.6	301150	1 AP003190	AP003190 Clostridi
14	119	11.4	10307	1 AE013241	AE013241 Mechanosa
15	118.5	11.4	10507	1 U32721	U32721 Haemophilus
16	117.5	11.3	3141	10 AF363030	AF363030 Mus muscu
17	117.5	11.3	3337	9 BC014434	BC014434 Homo sapi
18	117.5	11.3	4039	10 RNT76557	RNT76557 Rattus norv
19	117.5	11.3	5505	9 HSM804396	HSM804396 Neisseria
20	117.5	11.3	10612	1 AE002479	AE002479 Neisseria
21	117.5	11.3	329861	1 NMA52491	NMA52491 Neisseria
22	117.5	11.3	349980	6 AX044032	AX044032 Sequence
23	116	11.1	3161	3 AF145622	AF145622 Drosophi
24	116	11.1	3360	3 AF217788	AF217788 Drosophi
25	115.5	11.1	10671	1 AE007842	AE007842 Clostridi
26	115	11.0	10924	1 AE004238	AE004238 Vibrio ch
27	114.5	11.0	13991	1 AE010832	AE010832 Mechanosa
28	114	10.9	3246	8 LES312094	LES312094 Lycopersi
29	113.5	10.9	279160	1 AP000995	AP000995 Thermopla
30	113	10.8	2900	8 SCYB1084C	SCYB1084C
31	113	10.8	78600	8 SCDCNCH2	SCDCNCH2
32	112.5	10.8	297750	8 PHY17720	PHY17720 Pectunia hyb
33	112	10.7	3559	8 TACID4	TACID4
34	111.5	10.7	12055	1 AE000798	AE000798 Mechanoba
35	111	10.6	15428	1 AE001815	AE001815 Thermocog
36	111	10.6	17148	1 AE004798	AE004798 Pseudomon
37	111	10.6	286550	1 SME591785	SME591785 Sinorhizo
38	109.5	10.5	348550	1 AP003596	AP003596 Neococ sp
39	109	10.5	333800	1 SME591792	SME591792 Sinorhizo
40	109	10.5	346294	1 AP002999	AP002999 Mesorhizo
41	107.5	10.3	10784	1 AE010806	AE010806 Mechanosa
42	107.5	10.3	14876	1 AE009671	AE009671 Brucella
43	107	10.3	10182	1 AE011837	AE011837 Xanthomon
44	107	10.3	13968	1 AE010676	AE010676 Mechanosa
45	106.5	10.2	4241	1 AF269535	AF269535 Staphyloc

ALIGNMENTS

RESULT 1

log
seq file
pseud

AX081151 LOCUS AX081151 612 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 3 from Patent WO0109330.
ACCESSION AX081151
VERSION AX081151.1 GI:13170044
KEYWORDS Moraxella catarrhalis.
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 612)
AUTHORS Thomard,J.S.
TITLE Moraxella catarrhalis antigen basp121
JOURNAL Patent: WO 0109330-A 3 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
source location/Qualifiers
1..612
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 199 a 116 c 131 g 166 t
ORIGIN
Alignment Scores:
Pred. No.: 4.25e-98 Length: 612
Score: 1043.00 Matches: 204
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-069-544-2 (1-204) x AX081151 (1-612)
QY 1 MetLysIleArgValIleSTPProMetValMetAlaMetGlyLeuValLeuSerAlaCys 20
DB 1 ATGAATAATCCGAGTAAGTGGCTATGGTAATGGAGGGCTGCTTTAAAGTCCCTGT 60
QY 21 GlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIle 40
DB 61 CAAAGTACACCAATCCGCTTAAATATCCGCAATGGCAATCCGAATCCGAATCCGAATTT 120
QY 41 AlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAla 60
DB 121 GCAATCTCGCTACTGACATGGGTAAAGCTTGATCAAGCAACAGCAGTTGGATGCTGCT 180
QY 61 LeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAla 80
DB 181 TTATCGGCTGATCGGACATGGCTGCTATCGCAGCTTGCAGAAAGCTTTATCAAGCT 240
QY 81 SerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGluLysAlaIleGluLeu 100
DB 241 TCGAGAGATGCCATCCACCAACCAAGCTCAAGCTTGTGAAAAGCGCATTCAGACTA 300
QY 101 AsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAsp 120
DB 301 AATCTTAAAGATATGCAAAAGTTATATGATTAATGATTATTTGGTGCAGATGGGGAC 360
QY 121 LeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGlnGlyArg 140
DB 361 TTGTCAGGTGGCTGATTTATTTGATTAACCTTACCGACGATCGGTTATGAAGGGGCT 420
QY 141 ValValAlaIleGluAsnMetAlaTyrIleTyrThrHisGlnTyrGlnAlaAlaLysSer 160
DB 421 GTGGAGCCATCGAAAATATATGCAATATTTATATCATCAATATGAAGCGCAATCA 480
QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
DB 481 CCAACAAAGATGATCAATATATGCAATCAAGCACTTGAGCGGCTTATTTCAAGGC 540
QY 181 ThrGlnHisAspGlnIleLysLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeu 200
DB 541 ACTCAACATGATGAATCAAAAGCTTATGACAAATATTAAGTATTAATAATTAATA 600
QY 201 SerAspTyrLys 204

DB 601 AGTATTTTATAA 612
RESULT 2
LOCUS AX081149 615 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 1 from Patent WO0109330.
ACCESSION AX081149
VERSION AX081149.1 GI:13170043
KEYWORDS Moraxella catarrhalis.
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 615)
AUTHORS Thomard,J.S.
TITLE Moraxella catarrhalis antigen basp121
JOURNAL Patent: WO 0109330-A 1 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
source location/Qualifiers
1..615
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 200 a 116 c 132 g 167 t
ORIGIN
Alignment Scores:
Pred. No.: 4.27e-98 Length: 615
Score: 1043.00 Matches: 204
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-069-544-2 (1-204) x AX081149 (1-615)
QY 1 MetLysIleArgValIleSTPProMetValMetAlaMetGlyLeuValLeuSerAlaCys 20
DB 1 ATGAATAATCCGAGTAAGTGGCTATGGTAATGGAGGGCTGCTTTAAAGTCCCTGT 60
QY 21 GlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIle 40
DB 61 CAAAGTACACCAATCCGCTTAAATATCCGCAATGGCAATCCGAATCCGAATTT 120
QY 41 AlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAla 60
DB 121 GCAATCTCGCTACTGACATGGGTAAAGCTTGATCAAGCAACAGCAGTTGGATGCTGCT 180
QY 61 LeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAla 80
DB 181 TTATCGGCTGATCGGACATGGCTGCTATCGCAGCTTGCAGAAAGCTTTATCAAGCT 240
QY 81 SerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGluLysAlaIleGluLeu 100
DB 241 TCGAGAGATGCCATCCACCAACCAAGCTCAAGCTTGTGAAAAGCGCATTCAGACTA 300
QY 101 AsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAsp 120
DB 301 AATCTTAAAGATATGCAAAAGTTATATGATTAATGATTATTTGGTGCAGATGGGGAC 360
QY 121 LeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGlnGlyArg 140
DB 361 TTGTCAGGTGGCTGATTTATTTGATTAACCTTACCGACGATCGGTTATGAAGGGCT 420
QY 141 ValValAlaIleGluAsnMetAlaTyrIleTyrThrHisGlnTyrGlnAlaAlaLysSer 160
DB 421 GTGGAGCCATCGAAAATATATGCAATATTTATATCATCAATATGAAGCGCAATCA 480
QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
DB 481 CCAACAAAGATGATCAATATATGCAATCAAGCACTTGAGCGGCTTATTTCAAGGC 540
QY 181 ThrGlnHisAspGlnIleLysLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeu 200

|||||
Db 541 ACTCAACATGATGAATCAAAAAGCTTATGACAAATTATTAAGTATTAATTAATTA 600
Oy 201 SerAspTyrLys 204
Db 601 AGTGATTATTA 612
RESULT 3
AX067448 33140 bp DNA linear PAT 24-JAN-2001
LOCUS Sequence 23 from Patent WO0078968.
DEFINITION AX067448
ACCESSION AX067448.1 GI:12545068
VERSION
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
REFERENCE 1 (bases 1 to 33140)
AUTHORS Lagace, R.E., Patterson, C. and Berg, K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 23 28-DEC-2000;
Incye Genomics, Inc. (US)
FEATURES
source 1. .33140
/db_xref="taxon:480"
118
unsure /note="a or g or c or t, unknown, or other"
BASE COUNT 9472 a 6496 c 7384 g 9787 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 7,12e-96 Length: 33140
Score: 1043.00 Matches: 204
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-069-544-2 (1-204) x AX067448 (1-33140)
Oy 1 MetLysIleArgValYstrPProMetValMetAlaMetGlyLeuValLeuSerAlaCys 20
Db 21062 ATGAATATCGAGTAAGTGGCTGATGATGCGATGGGGCTGTTTAAGTGGCTGT 21121
Oy 21 GluSerThrProIleProPolysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIle 40
Db 21122 CAAAGTACACCAATACCGCTAAATAATCCGCAATGGCACAAATCCGAACTT 21181
Oy 41 AlaIleSerLeuAspMetGlyValLeuAspGlnAlaValGlnIleLeuAspAlaAla 60
Db 21182 GCAATCTCGTACTGACATGGGATGAGCTGATCAAGCAACGACGTTGATGCTGT 21241
Oy 61 LeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaValTyrGlnAla 80
Db 21242 TTATCGGCTGATGGGAGTGTGACCTGCTATCGCCTGGCAAGGTTTATCAAGCT 21301
Oy 81 SerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGluValAlaIleGluLeu 100
Db 21302 TCCGAGGATGCCACTCAACCAACCAAGCTCAACGCTGTTGTAAGGCGATTGAACCTA 21361
Oy 101 AsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAsp 120
Db 21362 AATCCCTTAAGATATGCAAGATATATGATATGATTTATTTGTCAGATGGGGAC 21421
Oy 121 LeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArg 140
Db 21422 TTGTCAGGTGCGTTGATTTATTTGATTAACCTAAGCGACCATCGTTATGAAGGCGCT 21481
Oy 141 ValValAlaIleGluAsnMetAlaTyrIleTyrThrHisGlnTyrGluAlaAlaLysSer 160
Db 21482 GTGGTACCATCGAAATATATGCGATATATATATATCATCATATATGAAGCTGCCAATCA 21541

Oy 161 ProThrLysAspAspTyrZnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
Db 21542 CCACAAAAGATGACATTAATAAGCCAAATACGACTTGAGGCTCGTTAATTTACGC 21601
Oy 181 ThrGlnHisAspGlnIleLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeu 200
Db 21602 ACTCAACATGATGAATCAAAAAGCTTATGACAAATTATTAAGTATTAATTAATTA 21661
Oy 201 SerAspTyrLys 204
Db 21662 AGTGATTATTA 21673
RESULT 4
AF074396 5628 bp DNA linear BCF 17-DEC-1998
LOCUS AF074396
DEFINITION Desulfotomaculum thermocisternum UDP-acetylglucosamine-
1-carboxyvinyltransferase (murA) gene, partial cds; yyda,
ferridoxin (fdx), dissimilatory sulfite reductase subunit A (dsrA),
dissimilatory sulfite reductase subunit B (dsrB), and dsrD genes,
complete cds; and unknown gene.
ACCESSION AF074396
VERSION AF074396.1 GI:4028016
KEYWORDS
SOURCE Desulfotomaculum thermocisternum.
ORGANISM Desulfotomaculum thermocisternum.
Bacteria; Firmicutes; Clostridia; Clostridiales; Peptococcaceae;
Desulfotomaculum.
REFERENCE 1 (bases 1 to 5628)
AUTHORS Larsen, O., Lien, T. and Birkeland, N.-K.
TITLE Dissimilatory sulfite reductase from Archaeoglobus profundus and
Desulfotomaculum thermocisternum: phylogenetic and structural
implications from gene sequences
JOURNAL Extremophiles (1998) In press
2 (bases 1 to 5628)
Larsen, O.
DIRECT SUBMISSION
SUBMITTED (25-JUN-1998) Department of Microbiology, University of
Bergen, Jahnebakken 5, N-5020 Bergen, Norway
FEATURES
source 1. .5628
/organism="Desulfotomaculum thermocisternum"
/strain="S190"
/db_xref="taxon:42471"
<1. .732
/gene="murA"
<1. .732
/gene="murA"
/note="similar to Bacillus subtilis UDP-acetylglucosamine
1-carboxyvinyltransferase"
/codon_start=1
/transl_table=11
/product="UDP-acetylglucosamine 1-carboxyvinyltransferase"
/protein_id="AAC96104.1"
/db_xref="GI:4028017"
/translation="DLANFLNCIGARVRGAGTDVIRKIEGVDISGGLRYAVIPDRIEA
GPMVAAATPBGVDLENYIPBILATIKAEAGVEGSEEDRVRVRSAGLSPDI
KTMYPFGPTDQSQMALLSTVPGTSVVENIPERFRVABELKMGRIKYEGLA
VVEGVERLQACVRATDLRAGALVVAAGLMAEGETRIDNVQYIDRGYFNLDEKLRMLG
ARIWRAPYDEKTPADDNAVRKSAI"
755. .1234
/gene="yyda"
755. .1234
/gene="yyda"
/note="similar to Bacillus subtilis yyda"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC96105.1"
/db_xref="GI:4028018"
/translation="WHITLVCVGLKEDFLLAQKEYLKRSPYARIELVEYREESLP
ASGRAGLAAVAKSGTRLLERI PNCVITADLRKQMLSSKEFAGYLAELGLRGNK
LCFIIGSGAGLADVLARASLRISFSKFTFPHQWRILILEOIYRAFKIIRGKYL"

```

gene      1494..1679
          /gene="fdx"
          /db_xref="GI:4028019"
          /note="similar to Fe4S4-ferredoxins"
          /codon_start=1
          /transl_table=11
          /product="ferredoxin"
          /protein_id="AAC96106.1"
          /db_xref="GI:4028019"
          /translation="MYVVTVDREKCEGCGACDSCPSVLEMKDKAFVANPDDCLGC
          ETCVCPGSAVLEER"
          1802..3217
          /gene="dsrA"
          /db_xref="GI:4028021"
          /note="similar to Desulfovibrio vulgaris dissimilatory
          sulfite reductase subunit A"
          /codon_start=1
          /transl_table=11
          /product="dissimilatory sulfite reductase subunit A"
          /protein_id="AAC96107.1"
          /db_xref="GI:4028020"
          /translation="MFEPRKPKALKYELRIYDAELNNYTEBELKEFKIKHPKY
          BELKGPMPFVADAKOELARLADDDMMIDRYVEDLGGLELSPNDGETHMKG
          GIVGVPFGVGGVGRSDVDPEKPGIAHPIHRYNOPASRKYKDFLALADLMEYRG
          SGLEPMGSGSDITLLITTTTQLEPIFYDMHELDLGGSGSLRTPSCIGTARCE
          WACTDTEMCTEMTHITODELHPAFYKFKFKFDGCPNCCVATADADAFITGTRD
          EIRIDQAVKAVIAGDIIVPRGAHGRDWGKFKDIQKAVIDRCPSQCMWMDGKVIDN
          KEVCVCHNCINMPQALPGKDTGVSILVGAKAPILGGAQIQAQIVPMAAEPRYDQ
          KEVERIMEFMEEGKNERVYGGELIQRVGLPKFLEVIQLPAPQIMIKAPRNPYIFWK
          EEDVPGFTRIDKEFRARHKR"
          3238..4425
          /gene="dsrB"
          /db_xref="GI:4028021"
          /note="similar to Desulfovibrio vulgaris dissimilatory
          sulfite reductase subunit B"
          /codon_start=1
          /transl_table=11
          /product="dissimilatory sulfite reductase subunit B"
          /protein_id="AAC96108.1"
          /db_xref="GI:4028021"
          /translation="MALSTGRFGKFDPPONTKORITDIPGRHYWENPPIIOKNYGM
          LYHELRPGVIVHVSFTGDKYTVRCGSPRLMTAOHIREICDIDAKYCDQVFRRTTN
          NVERIVDSQERVEALKKDLMKRFKRVSSYKPIGCTGASITNIYHOGYVHCPTPAD
          ASSLVKAVMDELFDFTGMILPAKVRVAVACCLMCAVHSDIALGIRKRPVIDH
          EVIDQVCELPLVISSCPGLAISPDKTEGSRSLKIYERKCMFCGNCYTCPLPIADK
          EGDQVAILVGGKISNAVSEPKSVVAVLPPNRPPEPVQIVKILVEYVADARHG
          ERLGDWVERIGMEKFEKTIPIFTEHLIDVRLAYDFTSTQKFTFAAMVCKAAG
          GID"
          4446..4742
          /gene="dsrD"
          /db_xref="GI:4028021"
          /note="similar to Desulfovibrio vulgaris dsrD and
          Archaeoglobus fulgidus dsrD"
          /codon_start=1
          /transl_table=11
          /product="unknown"
          /protein_id="AAC96109.1"
          /db_xref="GI:4028022"
          /translation="MKTPAOVTSYIELERYGVSEIEKKAIIIFPANSKKTRFYFD
          MEKAVOKETIPAKAREIKKASELVNGLTILFTSTTMYGLKERKASDEPQ"
          4856..5530
          /note="similar to Methanobacterium thermoautotrophicum
          O-linked glucan transferase"
          /codon_start=1
          /transl_table=11
          /product="unknown"
          /protein_id="AAC96110.1"
          /db_xref="GI:4028023"
          /translation="MGSGGGSNAFVERQVNNLQSNPGCATAKYNI/GVLMLEQGRLEBA

```

```

KTLFEALADGTBMPEAVYVNLGITVEKQCNLEKVECNRLAVOIEPRVAGYANLGFV
YLQMEKPDPAIEVLQKALIELNPRILQAMNANAYLQKDLQBALETNKKMLIADFE
SLGHNNLAVYIYILKGMNRAEHLKQALELGEBAHPEFLKEMEPYIANTHHDGNGHCP
STSSDA"
BASE COUNT      1397 a      1330 c      1629 g      1272 t
ORIGIN
Alignment Scores:
Pred. No.:      1.81e-05      Length:      5628
Score:          151.50      Matches:      48
Percent Similarity: 40.78%      Conservative: 36
Best Local Similarity: 23.30%      Mismatches: 69
Query Match:     14.53%      Indels:      53
DB:              1      Gaps:      5
US-10-069-544-2 (1-204) x AF074396 (1-5628)
QY      28  LysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuAspMet 47
          :::::|||||
          4910 CAGAGCAGACCCGCGCTGCGGCAGCCAGCAATCATCTGGAGTTAGCTGATGAACAG 4969
          Db
          QY      48  GlyLeuLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaSerAlaAspArgGlnPhe 67
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          4970 GGCCGCTGAGAGGAGCAAAACACTTTTGAAGAGCCATACCGATGGAGCGCGCARG 5029
          Db
          QY      68  AlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAspAlaThrIsgln 87
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5030 TTGGAAGCTTATGTTTAACTCGGTTATATTATTATTAAACAGGGGAACCTGGAAAGG 5089
          Db
          QY      88  ThrLysAlaGlnArgLeu-----IleTyrPheAspLysPro 131
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5090 GAGAGATGTAACCGCTGCGGGGTCAGATCGAACCCGTTAGCCCGCGCTATGCAAC 5149
          Db
          QY      94  -----PheGlnLys 96
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5150 CTGGCTTTGTCTACTCGATGAGTGAAGACCCGAGCAACCTTAAGTTCTGCAAAA 5209
          Db
          QY      97  AlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuVal 116
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5210 GCATTTGATGATGATCCAGATATATTCAGCCTGTGACCACTGGCAACGCTTACCTC 5269
          Db
          QY      117  GlnMetGlyAspLeuSerGlyAlaLeu-----IleTyrPheAspLysPro 131
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5270 CAGAAAGTGATCTGATCGCGGCGCTGAGACCAATCGAAGTGTGGAATTGCAACC 5329
          Db
          QY      132  SerArgAlaIleGlyTyrGlnGlyArgValAlaIleGlnAsnMetAlaTyrIleTyr 151
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5330 GACTTTTCGCTGGCCAC-----ATAACCTGGCTTACGTTAT 5368
          Db
          QY      152  TyrHisGlnTyrGlnAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSer 171
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5369 TAC-----CTGAAAGGGGACATGAACAGCGCGCGAG 5401
          Db
          QY      172  AlaLeuGlnArgAlaLeuIleSerGlyThrGln--HisAspGluIleLysLysSerTyr 190
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5402 CACCTGAGCAGCGCTGGAATCGGTTTGAAGCCCATCCGAATTTTAAAGAAATG 5461
          Db
          QY      191  AspLysLeuLeuSerAsp 196
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
          5462 GAACCATATCTCAATAC 5479
          Db
RESULT 5
AE013738      11161 bp      DNA      linear      BCT 26-JUL-2002
LOCUS        Versinia pestis KIM section 138 of 415 of the complete genome.
DEFINITION   AE013738 AB009952
ACCESSION   AE013738.1 GI:21958120
VERSION      AE013738.1 GI:21958120
KEYWORDS
SOURCE
ORGANISM     Versinia pestis KIM.
              Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Versinia.
REFERENCE    1 (bases 1 to 1161)

```

AUTHORS Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G.F., Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C., Peterson, J.D., Lindler, L.E., Brubaker, R.R., Plana, G.V., Straley, S.C., McDonough, K.A., Nilles, M.L., Watson, J.S., Blattner, F.R., and Perry, R.D.
Genome Sequence of *Yersinia pestis* KIM J. Bacteriol. 184 (16), 4601-4611 (2002)

TITLE 2 (bases 1 to 1161)
JOURNAL
PUBMED
12142430

AUTHORS Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G.F., Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., Schwartz, D.C., Peterson, J.D., Lindler, L.E., Brubaker, R.R., Plana, G.V., Straley, S.C., McDonough, K.A., Nilles, M.L., Watson, J.S., Blattner, F.R., and Perry, R.D.
Direct Submission
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers

FEATURES
source
1. 11161
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
108..857
/gene="Y1351"
108..857
/function="cell exterior constituents: surface structures"
/note="residues 10 to 227 of 249 are 37.15 pct identical to residues 16 to 229 of 237 from GenPept : >gb|AF94766.1| (AE004238) fimbrial biogenesis and twitching motility protein, putative [Vibrio cholerae]"
/codon_start=1
/transl_table=11
/product="putative fimbrial biogenesis protein"
/protein_id="AA084924.1"
/db_xref="GI:21958121"
/translation="MKLTLMRVCLVSVLTGSGRPPTNSQAVAGOTRLQGLAVL AAGDLTPARKLKEAVADPDYRTQGMAPFORIGENSAEQRYSQAKLPGNLT VLNNGAFSLGIGYVSAQOQFSAALPLPGYVORLIDENAGCFLPANDOKKAVYL SRALKYDPDGEPLIAEQRHREGNRQAQQLIDYQHTLPASAESLMQITFALALA GRDVSVRYSKQLARSFPQSKQYOHFLANEY"
847..1884
/gene="Y1352"
847..1884
/gene="Y1352"
/function="putative membrane"
/note="residues 1 to 345 of 345 are 57.14 pct identical to residues 1 to 332 of 337 from E. coli K12 : B2516; residues 1 to 345 of 345 are 59.77 pct identical to residues 1 to 329 of 334 from GenPept : >gb|AF121418.1| (AE008814) paral putative membrane protein [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AA084925.1"
/db_xref="GI:21958122"
/translation="MNTESQDQCTTETPEVRLRQARESLGTLQQTVAERLCLKVSTI RDIIEEDNAQANLASTFPRGYIRSYAKVHLPEDELPILEKQAPVRKAVAPQSGSL GKHKRKGWLMSTWLVVLGTGAMWQNHQAQAEIANNVDQSSQQLSQNGQGPVPLTQNSDAIAPTDAPAVANGQPVPLRNHSGSAITNSATTSVPKTSPEVDA NTNTTMOGEAASAAVSPQVPOGMPGTQPLPTADAGVSGASASVAGALVWNPETDCA WLOVVDATGKTLFSGIOKGAVNLNAGKAPYKLTIGAPALTTISYQNPVDSLKPTIKA NRYARLTVCVE"
1946..3073
/gene="gcpB"
/note="Y1353"
1946..3073
/gene="gcpB"
/note="residues 1 to 372 of 375 are 86.82 pct identical to residues 1 to 371 of 372 from E. coli K12 : B2515"
/codon_start=1
/transl_table=11

gene
CDS
/product="hypothetical protein"
/protein_id="AA084926.1"
/db_xref="GI:21958123"
/translation="MNGSPITRRKSTRIVYGVKVPVIGDGAPIAVQSMITKTVDVAT VAOIKALERVGVDIVRVSIPMDAAEFKLIRKQOSTVPPVADIHPRIALKAEYGV DCIRINPQNGVGNESRIREVACADRYNIPIRIGINQSELEKIOERYGEPTEALLES AMHVDILDRNFQDFKVSADSVLAANSYLLAKQINNPHLIGITFAGAGRSQSV KSAIGLLSEIGDGLTRISLAADVEKVGDFDLKSLRIRARGINFACPTGSGRO EPRVGTVALEQRLEDLITPMQVSIIGCVNNGEGALVSTICGVARRHSGREYECV ROKERFDNKAMIDQLEAKTRAKASIIIDANNRIYINQDNK"
3282..4556
/gene="hisS"
/note="Y1354"
3282..4556
/gene="hisS"
/function="enzyme; aminoacyl tRNA synthetases, tRNA modification"
/note="residues 1 to 424 of 424 are 81.83 pct identical to residues 1 to 424 of 424 from E. coli K12 : B2514; residues 1 to 424 of 424 are 81.83 pct identical to residues 1 to 424 of 424 from GenPept : >gb|AF121416.1| (AE008814) histidine tRNA synthetase [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="histidine tRNA synthetase"
/protein_id="AA084927.1"
/db_xref="GI:21958124"
/translation="MAKNIOAIRGNMDYLPADTAIMQRIEISILKQVSGYSEIRMP IYEOPLFRAIGEIVDVEKEMETFPDNRNGBSLTPRPGTAGCAGAGIEHGLVYQE ORLYIGPFRERYERPOKGRYRPOHGAEVFGLPGPDIAEILILFARMRALGIEFH VLELNSIGSLIARADREARVALFEOHVEVLDDECRKMSNPLEPLDSKNDVOOL LDDAPKLSYLDIBESKQHPAGACELLDKASITPTVERLVRCGLDYNNRYFEVYTHSL GAGTVCAGGRYDGLVEQDGRTPAVGAGUEBRVLLVQANAFQVPATVDAYV SSGEQASQASAMLAESLRDALPLTKIMTYGQGNVKKQFTRADKQARVALMIGSEV AAQOVVVKDLRNGEQTTLAQADVAAALALMLG"
4570..5205
/gene="Y1355"
4570..5205
/gene="Y1355"
/note="residues 1 to 210 of 211 are 57.14 pct identical to residues 1 to 205 of 206 from E. coli K12 : B2513; residues 1 to 210 of 211 are 59.04 pct identical to residues 1 to 205 of 206 from GenPept : >gb|AF121415.1| (AE008814) putative inner membrane protein [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084928.1"
/db_xref="GI:21958125"
/translation="MEVYTTENDQVDARVRFSENGKALAVGVIGALVGNRWYOG NQNTAMTEASSAYQKYSTALSTNSAGNVASDIALAEKFSQSNKNYGVYVGLIELAD FVEQNAFDAQOOLSQLGQTKDENLISLRILAVQQLKKPDPALTKLIDVQDGG WTAQAQDVARDGDLKSGDTAGARAASAKIESNASQALQALRMKLNINS"
5217..6398
/gene="Y1356"
5217..6398
/gene="Y1356"
/function="putative"
/note="residues 1 to 393 of 393 are 76.33 pct identical to residues 1 to 392 of 392 from E. coli K12 : B2512; residues 1 to 393 of 393 are 76.33 pct identical to residues 1 to 392 of 392 from GenPept : >emb|CAD02723.1| (AF162725) putative lipoprotein [Salmonella enterica subsp. enterica serovar Typhi]"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AA084929.1"
/db_xref="GI:21958126"
/translation="MQLRKLTVGLVSVALLSGCSLFSNEEDVVTWSPLPKYENQSPF TKWSTSVSGVGDYYSHLRPAWQGTSVFAADRKKGLVYKAMDADVTGQIYNOTDISEKTN

misc_feature
/translation="MPKPKQKALKOKGKMLKGMKWFNFYSAILLFLVPLVIVPL
SFGAGSLSTYPMAGFSLRWYQTFNSDEMGALGNSLLIAPLATLGTGVAAGL
VFEGRFKALPMAMLIISPMVAIVVIGMFEFFARLSLNSYGLVLAALGVPV
VITVAVKISDOLSRRAASLIGASOMRVRKRTVLPILAGVFGALFAVAVSDEV
VTFLASPROGRTLPIMFAGIRENLDTPTAAASLWYGASLLLIWELRRSRERL
HHPSP"
complement(1292.357)
/gene="YPO2841"
/note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(1376.597)
misc_feature
/note="Pfam match to entry PF00528 BPD_transp,
Binding-protein-dependent transport systems inner membrane
component, score 59.50, E-value 7.3e-14"
complement(454.519)
misc_feature
/gene="YPO2841"
/note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(508.594)
misc_feature
/gene="YPO2841"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane comp. sign."
complement(583.648)
misc_feature
/gene="YPO2841"
/note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(694.759)
misc_feature
/gene="YPO2841"
/note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(787.852)
misc_feature
/gene="YPO2841"
/note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(967.1032)
misc_feature
/gene="YPO2841"
/note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(1157.2422)
misc_feature
/gene="YPO2842"
complement(1157.2422)
gene
/gene="YPO2842"
/note="Similarity to the N-terminus of Agrobacterium
tumefaciens mannopine permease MOPC TR:Q44382
(EMBL:AF242881) (294 aa) fasta scores: E(): 2.7e-29, 41.3%
id in 213 aa, and to the full length Pseudomonas
aeruginosa probable ABC transporter permease PA0605
TR:AG03994 (EMBL:AE004497) (415 aa) fasta scores: E(): 0,
54.7% id in 397 aa"
/codon_start=1
/transl_table=1
/product="putative ABC transporter permease"
/protein_id="CAC92094.1"
/db_xref="GI:15980812"
/db_xref="SPTREMBL:Q8ZCW6"
/translation="MSQTEWYTKVADEROQSGPILRQOLRKVKAAYTRSLITPL
LPLVYSFPIATLIGKSIDNPETLRAMVAVLANWGDALPDEAREALITDR
MARNSGIMLSIAKMSYEDNRYSLSINSTIRKVPAGEVREALISALPLGELSTWOS
LQRAQPLTRLYKLPASDRVAVESGQIKLPADQALYVNLRTLMAAVTLLCT
LGYPLAVMLKQPTGRANLLMLVLPEFTSLVTRTSMVLLQSSGLINQSLIGSLT
IEOPLAVFNRAGYISMTILLPEYVLPVAVMKISPNYVRAAISLIGAPPAER
VYVQFNGTAGALLVFMAGIGYITRIPALIGSGQDMVSYPVAFPTNTMMGMMAA
LGSOLLITIVLLVYVIRVTRQAEFTAIH"
complement(1196.1261)
misc_feature
/gene="YPO2842"
/note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1277.1498)
misc_feature
/gene="YPO2842"
/note="Pfam match to entry PF00528 BPD_transp,
Binding-protein-dependent transport systems inner membrane
component, score 40.70, E-value 3.2e-08"

misc_feature
complement(1328.1393)
/gene="YPO2842"
/note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1409.1495)
misc_feature
/gene="YPO2842"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane comp. sign."
complement(1493.1558)
misc_feature
/gene="YPO2842"
/note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1652.1717)
misc_feature
/gene="YPO2842"
/note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1739.1804)
misc_feature
/gene="YPO2842"
/note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(2246.2311)
misc_feature
/gene="YPO2842"
/note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(2469.3497)
misc_feature
/gene="YPO2843"
complement(2469.3497)
CDS
/gene="YPO2843"
/note="Similar to Agrobacterium radiobacter agropine
permease protein AgCB TR:Q9R701 (EMBL:AF242881) (364 aa)
fasta scores: E(): 9.9e-26, 30.3% id in 320 aa, and to
Pseudomonas aeruginosa hypothetical protein PA0222
TR:AG03611 (EMBL:AE004460) (352 aa) fasta scores: E(): 0,
56.1% id in 342 aa"
/codon_start=1
/transl_table=1
/product="putative transport system periplasmic protein"
/protein_id="CAC92095.1"
/db_xref="GI:15980813"
/db_xref="SPTREMBL:Q8ZCW5"
/translation="MPKMTLTLLAGVFGTHAETLVVSFGGLKNOQVAKFYKPF
/MAAGKTIAGEVNGEMARIRAMVETGQIGMDIVEEGELLRGCEGLFETLWSKL
GDERPFIKGTGVCQSGSLFPLSMMLIYDASKKPTVTSWADFNWVTFPGKSLRTA
KFMETALADADVKEDVRYLVATPEGVGVENKLDQKPNOMWESGAPLOMWLVS
DYVMSVYVARGVALKEGHDPRMTWTSITYMDSTIYKSGKKYLAQCFALFANLA
ENQKFAENIAGPTINATLMIDPEIAANLPTAPANLAEAPMNNQFVHSEDLQ
RNSWAK"
complement(3550.4839)
gene
/gene="YPO2844"
/note="gabt: goag"
complement(3550.4839)
CDS
/gene="YPO2844"
/BC_number="2.6.1.19"
/note="Similar to Escherichia coli 4-aminobutyrate
aminotransferase Goag SW:GOAG_ECOLI (P50457) (421 aa)
fasta scores: E(): 0, 65.7% id in 428 aa, and to
Pseudomonas aeruginosa 4-aminobutyrate aminotransferase
Gabt TR:AG03655 (EMBL:AE004465) (426 aa) fasta scores:
E(): 0, 53.0% id in 423 aa"
/codon_start=1
/transl_table=1
/product="putative aminobutyrate aminotransferase"
/protein_id="CAC92096.1"
/db_xref="GI:15980814"
/db_xref="SPTREMBL:Q8ZCW4"
/translation="MSHRELQRRKREKATPRGIGVLCDFYAVRAENATLWDEQREYID
PFIAGIATLNIQHRHKKVAVAQQDQFHTAYQVYVASYTLAEKINSIAPISDN
MTAAGSKTAPFTTGVEIAENAVKIJARAATGPGVYAGFAGHRTILAMALTGRAVP
YVVGQPPASITFHALYPMELYGVSVERAISVERLFDGDISPTVAAALFPIQOEG
GPNIAPEPVSAIURLCEBHGILLIADREVQTSFARTGKIFANEYVPTDVTVDITAKS
LGGMPISAVTGRADIMDAPLPQSGYGYAGNPILVAASLAVLDIIAEKRLCALITL
GAKLVDALEKQMSNAIVGIRARSGMAVEFNDVSGPSPBELTRAYORALIEGLT
LISGVHSNVIIFLYPLTIPDKQKQAMNITRLIAS"


```
Qy      168 AsnAlaIysSerAlaLeuGluArgAlaLeuIleSerGlyThrGlnHisAspGluIleLys 187
      |||||
Db      175374 AAGCTAAGAAATTTATGAGAAAGCC-----ATTAAA 175342
      |||||
Qy      188 LysSerTyrAspLysLeuLeuSerAspTyrLysLeuSerAspTyrLys 204
      |||||
Db      175341 AAGAAATTTGGTTATGCTTTTCATCTTCTTAATTTGGACAGTATATATAG 175291
      |||||

RESULT 8
AE000799/c 10065 bp DNA linear BCT 19-JUN-2002
LOCUS      Methanobacterium thermoautotrophicum f. Delta H
DEFINITION (section 5 of 148) of the complete genome.
ACCESSION  AE000799 AE000666
VERSION     AE000799.1 GI:2621112
KEYWORDS
SOURCE
ORGANISM   Methanothermobacter thermoautotrophicus str. Delta H.
            Methanothermobacter thermoautotrophicus str. Delta H
            Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriales;
            Methanobacteriaceae; Methanothermobacter.
REFERENCE
AUTHORS    Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J.,
            Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K.,
            Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pochier, B., Qiu, D.,
            Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R.,
            Jiwani, N., Carnuso, A., Bush, D. and Reeve, J.N.
            Complete genome sequence of Methanobacterium thermoautotrophicum
            deltaH: functional analysis and comparative genomics
            J. Bacteriol. 179 (22), 7135-7155 (1997)
            9371463
JOURNAL    2 (bases 1 to 10065)
MEDLINE    Smith, D.R.
PUBMED     Direct Submission
            Submitted (10-AUG-1997) Genomics and Technology Development, Genome
            Therapeutics Corporation, 100 Beaver Street, Waltham, MA
            02154-8448, USA

FEATURES
Source     location/Qualifiers
            1..10065
            /organism="Methanothermobacter thermoautotrophicus str.
            Delta H"
            /strain="Delta H"
            /db_xref="taxon:187420"
            /clone="MTH7"
            /note="synonym: Methanobacterium thermoautotrophicum str.
            deltaH"
            complement(106..486)
            /gene="MTH76"
            /complement(106..486)
            /gene="MTH76"
            /note="Function Code:14.00 - Unknown, ; similar to,
            gp:GI:e288521:GI:1752890, p(1)=0.997, pid=*"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AAB84582.1"
            /db_xref="GI:2621113"
            /translation="MSFTGKIVFANLSEVGVQRGRCDRNFTCYNABRCISPPS
            AISRVSVPNNKPTSOYONKEADQISLQIASNLIIPSONNLMDOTPHYTNKGD
            RFPKRLSKFSSRPPIILFSSMTK"
            complement(125..999)
            /note="741 bp direct repeat includes part of MTH77
            (unknown - paralog.fam. 20); 97% ID to interval
            41041-40302"
            complement(587..1032)
            /note="446 bp direct repeat includes part of MTH77
            (unknown - paralog.fam. 20); 85% ID to interval
            50966-50521"
            complement(603..769)
            /note="167 bp direct repeat includes part of MTH77
            (unknown - paralog.fam. 20); 83% ID to interval
            54315-54149"

repeat_region
complement(604..1032)
/note="429 bp direct repeat includes part of MTH77
(unknown - paralog.fam. 20); 80% ID to interval
46089-45658"
complement(604..1011)
/gene="MTH77"
complement(604..1011)
/gene="MTH77"
/note="Function Code:14.00 - Unknown, ; similar to,
pir:LN:A39318, p(1)=0.97, pid=*"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB84583.1"
/db_xref="GI:2621114"
/translation="MEGFGIGNMAAYATIIAYGINTIDDIKKFIIHQTCLEDF
HGILDIATVAGVAVNGPIGFSIVIAEGDLIRITHTLIPQYMKYISIRHWITGY
ELRSYGEDHCIIYIEIPKPDGSLRWDAAAYI"
complement(615..798)
/note="184 bp direct repeat includes part of MTH77
(unknown - paralog.fam. 20); 75% ID to interval
49855-49672"
complement(895..999)
/note="105 bp direct repeat includes part of MTH77
(unknown - paralog.fam. 20); 100% ID to interval
50071-49967"
complement(1142..1384)
/gene="MTH78"
complement(1142..1384)
/gene="MTH78"
/note="Function Code:14.00 - Unknown, ; similar to,
sp:LN:YKF3 YEAST, p(1)=0.95, pid=*"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB84584.1"
/db_xref="GI:2621115"
/translation="MMVKYIFPFSYPLSYFTALVVVSIRFPYISKLIVSYIVNM
FLIIVLIGDELKGTENETVRALGGLVILIGFGLA"
complement(1398..1901)
/gene="MTH79"
complement(1398..1901)
/gene="MTH79"
/note="Function Code:14.00 - Unknown, ; similar to,
gp:GI:e220736:GI:177280, p(1)=0.88, pid=*"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB84585.1"
/db_xref="GI:2621116"
/translation="MIVRLRLIGFTSMYLSINLPLFRLCRASLIGLNMAAYATIIA
YTGILNITDIIKKFIIHQTCIGGDIIDGILFWSLKESEGVTLASGIFASILL
AEGDLIVNRTNLTLPTEYMKYVSYHRTMHWGVLKLYVEGDCHIIIEIPKNDGTLR
WDEAVYI"
complement(1400..1561)
/note="162 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 78% ID to interval
45821-45660"
complement(1402..1561)
/note="160 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 80% ID to interval
54313-54154"
complement(1404..1473)
/note="70 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 84% ID to interval
53622-53556"
complement(1409..1592)
/note="184 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 75% ID to interval
49061-48878"
complement(1686..1808)
/note="123 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 93% ID to interval
46053-45931"

repeat_region
complement(604..1032)
/note="429 bp direct repeat includes part of MTH77
(unknown - paralog.fam. 20); 80% ID to interval
46089-45658"
complement(604..1011)
/gene="MTH77"
complement(604..1011)
/gene="MTH77"
/note="Function Code:14.00 - Unknown, ; similar to,
pir:LN:A39318, p(1)=0.97, pid=*"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB84583.1"
/db_xref="GI:2621114"
/translation="MEGFGIGNMAAYATIIAYGINTIDDIKKFIIHQTCLEDF
HGILDIATVAGVAVNGPIGFSIVIAEGDLIRITHTLIPQYMKYISIRHWITGY
ELRSYGEDHCIIYIEIPKPDGSLRWDAAAYI"
complement(615..798)
/note="184 bp direct repeat includes part of MTH77
(unknown - paralog.fam. 20); 75% ID to interval
49855-49672"
complement(895..999)
/note="105 bp direct repeat includes part of MTH77
(unknown - paralog.fam. 20); 100% ID to interval
50071-49967"
complement(1142..1384)
/gene="MTH78"
complement(1142..1384)
/gene="MTH78"
/note="Function Code:14.00 - Unknown, ; similar to,
sp:LN:YKF3 YEAST, p(1)=0.95, pid=*"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB84584.1"
/db_xref="GI:2621115"
/translation="MMVKYIFPFSYPLSYFTALVVVSIRFPYISKLIVSYIVNM
FLIIVLIGDELKGTENETVRALGGLVILIGFGLA"
complement(1398..1901)
/gene="MTH79"
complement(1398..1901)
/gene="MTH79"
/note="Function Code:14.00 - Unknown, ; similar to,
gp:GI:e220736:GI:177280, p(1)=0.88, pid=*"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB84585.1"
/db_xref="GI:2621116"
/translation="MIVRLRLIGFTSMYLSINLPLFRLCRASLIGLNMAAYATIIA
YTGILNITDIIKKFIIHQTCIGGDIIDGILFWSLKESEGVTLASGIFASILL
AEGDLIVNRTNLTLPTEYMKYVSYHRTMHWGVLKLYVEGDCHIIIEIPKNDGTLR
WDEAVYI"
complement(1400..1561)
/note="162 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 78% ID to interval
45821-45660"
complement(1402..1561)
/note="160 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 80% ID to interval
54313-54154"
complement(1404..1473)
/note="70 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 84% ID to interval
53622-53556"
complement(1409..1592)
/note="184 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 75% ID to interval
49061-48878"
complement(1686..1808)
/note="123 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 93% ID to interval
46053-45931"
```



```

repeat_region complement(1704.1937)
/note="234 bp direct repeat includes part of MTH80
(unknown - paralog.fam. 126); 98% ID to interval
41170-40937"
repeat_region complement(1704.1808)
/note="105 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 100% ID to interval
49262-49158"
repeat_region complement(1705.1808)
/note="104 bp direct repeat includes part of MTH79
(unknown - paralog.fam. 20); 98% ID to interval
50933-50830"
repeat_region complement(1766.1813)
/note="48 bp direct repeat includes part of MTH79 (unknown
- paralog.fam. 20); 90% ID to interval 43075-43028"
repeat_region complement(1854.1937)
/note="84 bp direct repeat includes part of MTH80 (unknown
- paralog.fam. 126); 88% ID to interval 46202-46120"
gene /gene="MTH80"
complement(1917.2162)
/note="Function Code:14.00 - Unknown; similar to,
gp:GI:244452:9170601, p()=0.29, pid=*"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAB84586.1"
/db_xref="GI:2621117"
/translation="MDYSLDDLYSLICDILFISIGIHENGLIPADIALLPLFIS
LFLWLPKLERDEIRMLVAITFVSSSSMMATL"
repeat_region complement(1970.2064)
/note="95 bp direct repeat includes part of MTH80 (unknown
- paralog.fam. 126); 80% ID to interval 45540-45446"
repeat_region complement(2019.2170)
/note="152 bp direct repeat includes part of MTH80
(unknown - paralog.fam. 126); 80% ID to interval
53541-53391"
repeat_region complement(2258.2703)
/note="446 bp direct repeat includes part of MTH81
(unknown - paralog.fam. 20); 85% ID to interval
49295-48850"
repeat_region complement(2260.3830)
/note="1571 bp direct repeat includes part of MTH83
(O-linked GlcNAc transferase - paralog.fam. 17); 93% ID to
interval 42184-40631"
repeat_region complement(2274.3052)
/note="779 bp direct repeat includes part of MTH81
(unknown - paralog.fam. 20); 80% ID to interval
54942-54149"
repeat_region complement(2275.3052)
/note="778 bp direct repeat includes part of MTH81
(unknown - paralog.fam. 20); 90% ID to interval
46441-45658"
gene /gene="MTH81"
complement(2275.2919)
/note="Function Code:14.00 - Unknown"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAB84587.1"
/db_xref="GI:2621118"
/translation="MIYAFMGAYCYHQGTQEMARELGQALISEKLDLSLDGNGTVNLE"
Alignment Scores:
Pred. No.: 0.00338 Length: 10065
Score: 132.50 Matches: 49
Percent Similarity: 46.70% Conservative: 36
Best Local Similarity: 26.92% Mismatches: 70
Query Match: 12.70% Indels: 27
DB: 1 Gaps: 7

```

```

US-10-069-544-2 (1-204) x AE000739 (1-10065)
QY 28 LysAsnAsnProGlnInleuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMet 47
Db 4071 AAAAACAATCCAAACCTTGAGAGACATGACACACAAAGAGATGCTCAAGAACCTT 4012
QY 48 GlyLeuLeuAspGlnAlaLysGlnGlnIleuAspAlaAlaLeuSerAlaAspArgGlnPhe 67
Db 4011 GCGAGATGATGATAGGACCTAGAAATGCTATGAAAAGCACTTAAATTAATCCAAAGTAT 3952
QY 68 AlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnuAspAlaThrHisGln 87
Db 3951 GCGAGAGACATGCAACAATAAGACATGATGCTC-----AAAGAACTGGAGATAT 3901
QY 88 ThrLeuAlaGlnIleuArgLeuPheGluValAlaIleGlnLeuAsnProLysAspMetGlnSer 107
Db 3900 GATGAGGACCTAGAAATGCTATGAAAAGCACTTAAACCCAAACTTGACAGATGCA 3841
QY 108 TyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyr 127
Db 3840 TGGTACAAATTAAGGCTCCGTTCTAATTACCTTAAAAAATACAAAAGCACTTAAATGT 3781
QY 128 PheAspLysPro-----SerArgAlaIleGlyTyrGlnGlyArg 140
Db 3780 TTCGAAAAAGCCATAGAGCTAATCCAAAATTTATAGAGCTGGGGAACCTAAAGCA--- 3724
QY 141 ValValAlaIleGlnuMetAlaTyrIleTyrHisGlnTyrGlnAlaLys--- 159
Db 3723 --ATAACATTAATACCTTAAG--ATAATA-----GAAAGCATTTAAATGT 3679
QY 160 -----SerProThrLysAspAspTyrAsnAsnAlaLysSerAla 172
Db 3678 TATGACAAAGTTCTCCAAATTAATCCAAAGATGATTAAGCTTGAAACAAGAGACTA 3619
QY 173 LeuGlnArgAlaLeuIleSerGlyThrGlnHisAspGlnIleLysLysSerTyrAspLys 192
Db 3618 GTTTCATGAACTT-----GGAGATATGATGAGTACATGAAATGCTATGAAAAA 3568
QY 193 LeuLeu 194
Db 3567 GCATTA 3562

RESULT 9
CEU77412 4097 bp mRNA linear INV 19-JUL-1997
LOCUS CEU77412
DEFINITION Caenorhabditis elegans O-linked GlcNAc transferase mRNA, complete
cds.
ACCESSION U77412
VERSION U77412.1 GI:2266991
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
TITLE
1 (bases 1 to 4097)
AUTHORS Lubas W.A., Frank D.W., Krause M. and Hanover J.A.
JOURNAL O-linked GlcNAc transferase is a conserved nucleocytoplasmic
protein containing tetrapeptide repeats
J. Biol. Chem. 272 (14), 9316-9324 (1997)
MEDLINE 97238870
PUBMED 9083068
REFERENCE 2 (bases 1 to 4097)
Lubas W.A., Frank D., Krause M. and Hanover J.A.
Direct Substitution
Submitted (05-NOV-1996) Laboratory of Cell Biochemistry and
Biology, NIDDK, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES
source
1. 4097
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IIIc"
1. 3456
CDS

```



```
gene
3859..4404
/gene="PM2008"
3859..4404
/gene="PM2008"
/codon_start=1
/transl_table=11
/product="unknown"
/db_xref="GI:12722450"
/translacion="MTTRVQVWIASISLPLSACVSGECKTDSHEAARIELGIG
YLSQHDPAKQNDKALSHAPNYLSHAVALYVYQKQGLDLARHAYEKAITLQKQ
GDVLMNNGAPLCOQGEPEQAKQPSQALNAPHYQADTYENLWMCASFADTVRYQO
NLQHLQPDENRRAKKFPPIIE"
4527..5483
/gene="PM2009"
4527..5483
/gene="PM2009"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK04093.1"
/db_xref="GI:12722451"
/translacion="MTATANNIEKVTLLGEQRFQAREALNLSIEDVSKKISLRPSIL
QLLETQIAQKSIPTATMGVRSYAFKLPSVWVAHAFGEQDHKNDLQKNARATA
VNHYSNHNHIGWLSLVLIVAGMTGLMMENYOSNARESLVQHYVDTTQSSVY
QSTPIASTALTDTQNAVLASVVIPTTQTTSSDTLPTTISPSDQDITADVLK
AEMKLDKSEKULSHETVTPSHHEIETVGTGCWISQNESGKILAQEYKQGDVL
TNEBGVYSKILIGAPANVKITTYGEDIPLKVDGRVAFKLPQ"
5510..6613
/gene="gcpe"
/ncore="PM2010"
5510..6613
/gene="gcpe"
/codon_start=1
/transl_table=11
/product="Gcpe"
/protein_id="AAK04094.1"
/db_xref="GI:12722452"
/translacion="MLAKPRIKRRESKTYGVNVPICGDAPFAVQSMTRTTRTDEAT
VAQIKALRYGADIVRSVPTMDAAEFKIKQOVNPLVADHFDRIALKAELGV
DCIRINFGNIGREDRIKAVVDCARDNNIPTRIGVNGSLKEDLOEKGEPTKAEYV
ALHVEILDLRLNPDQKVSVDVLAESYLLAKAIQPLHIGITEAGCRAAV
KSAVGLMLAIEGIDTLRISLADPVEIKVGFIDIKSIRISRGINFACPTGRO
EPFVIGTVNALREQLIEDIITPMDSIIGCVNPGELVSDLGVTGKNSGYVDGE
ROKERPDNDLNLQLEKIRAKVAADPKRRIT"
6643..7914
/gene="his5"
/ncore="PM2011"
6643..7914
/gene="his5"
/codon_start=1
/transl_table=11
/product="His5"
/protein_id="AAK04095.1"
/db_xref="GI:12722453"
/translacion="MAKTOAIRGMDCSPTESPLQWVSGKRVSVLQNGVSEVMP
IVESTPLPARAISEVDVASKEMITPRDNDLQTLREBGTAGCVRAIEHGMVINOEQ
RLMYGMPFERHPQKGRYQFHQAGVEVGIENPETDAELMLTALWKEIGIAIYH
TLQNLISGLSEAKNYSALVAFLOQHVLDLSEBEKREKRLRILDTKNOQLQEV
LNDAPKLAYLDDDSREHFAQLCALDVAIGIYEVNPLVGLDYVNTVEVNTSALG
AGGTVCGGGYDGLVEQLGATTVGAFMGLERLVLLVEVNOGSIPLPAVDIYVYV
QSGETTLAARFQALRETLRTELPLRVMTHGCGGNPKQFRADSGATLALVIGESVO
NOOVVVKHLIGLGEQTLALDDIVDIYIKNF"
7933..8544
/gene="PM2012"
7933..8544
/gene="PM2012"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK04096.1"
/db_xref="GI:12722454"
/translacion="MAVTAEBEOLNIEKITMKENYKTLIASYIATFGVFGNWNQ
S HOLNKIQOTSAQVQALYSTADPASKAQIOECPVQANSKTSYVALLEKAYEKK
EPAQAEQSLKQALQASTDNLSSITLRLASVQYQQAQAFMDLSLKQYKESWDSRK
NLLIGDILLAKGDKVAAKASYQALANNAFLEQWLOVNLNL"
```

```
gene
8749..9186
/gene="PM2013"
8749..9186
/gene="PM2013"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK04097.1"
/db_xref="GI:12722455"
/translacion="MTIRIPPPYIFLCAAWMKCPRIYQAFHTWLQGLFISGII
GSGSLMQFWQATTVNDPIRLKSSALVVSIGIYQSPNPMYLSIALLLLMGLMLASLS
AVLFLVYVFCITTOFQIKPEBEROLEFDADYLDYKQYRRML"
complement(9298..11268)
/gene="cpdb"
/ncore="PM2014"
complement(9298..11268)
/gene="cpdb"
/codon_start=1
/transl_table=11
/product="CpDB"
/protein_id="AAK04098.1"
/db_xref="GI:12722456"
/translacion="MDRRPELOGCSILTFTRAKYVATPONSASIRIATTDHSF
LUDPDYKAPPEKRFPTAASLITROAREKNSVUVNDGDLQGNPLADYQAKYK
EKGMPAIEIVLNAKRYDVTGLNHEFNYSGLKTLDAIAANPPIINANVKAQSGEP
YQPYVIOEKGVNDANGHLOMIRIGYGFPPQVPMWDKALHGHVLAHDIKTAEKYV
PELKAKGADIIIALAHTGPSDEPKYEGKASAFYLAADVIGIAVIFGSHRLFPNKEF
AKSPNDIEKGTGKVGESMAGYMGWNNISVLDLQHDGDKVNVNKGALRPIYNE
KMAIAPDPDENNALIAPVHDAITROVSPGIRAVDNMSYALVODDPTQIVNQA
RAYAENVKUNPELAGLPLVSKAPFKACGRKNDPTGYTEVDKGLTFPMAADLYLP
NTLVIVKVSQALKEWLECSAGMFKQIDSNQPOSLDWTGERTYNTFVIDGVQYQF
DITTPARYDQDKLINAOSEKRVINLYQKQPYDPKAEFLIATNNRYAASAKFPGTDS
HIVFAAPDENROILLANYISSEKTHQVNSADKWMRIAPITSKVLDIRFETSPEK
AAQFOERQRYPMVVOGVTDEVGPAIYQVDSLQ"
```

```
BASE COUNT 3480 a 1995 c 2500 g 3293 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0368 Length: 11268
Score: 123.00 Matches: 45
Percent Similarity: 41.42% Conservative: 25
Best Local Similarity: 26.63% Mismatches: 77
Query Match: 11.79% Indels: 22
DB: 1 Gaps: 5
US-10-069-544-2 (1-204) x AE006237 (1-11268)
```

```
QY 15 LeuValLeuSerAlaCysGlnSerThrProIleProPolysAsnAsnPro----- 31
Db 3901 CTTTTCCTTTCGCGCTGT-----GTGAGTCAAGAGGTAACAAAGCTTGAT 3948
QY 32 -----GlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuSmetGlyLys 49
Db 3949 TCACATGAGCAGCAAAAGCCAGAAATTGATTGAGTTAGGCTATTATCTCAGCATGAT 4008
QY 50 LeuAspGlnAlaLysGlnLeuLeuAspAlaAlaLeuSerAlaAspArgLnpheAlaPro 69
Db 4009 TTTGCCGACGCAAGCAAAATTTGATTAAGCTTATCCCATGCCCAATTAATTAAT 4068
QY 70 AlaTyArgThrLeuAlaLysValTyArgIlnalaserGlnAspAlaThrHisGlnThrLys 89
Db 4069 TCTCATGCGGTCTTTCGCTTTCGCTATCTACCAAAAAA-----GTCATCTTGATTTA 4119
QY 90 AlaGlnArgLeuPheGlnLysAlaIleGlnLysLeuAsnProLysAspMetGlnSerTyrmec 109
Db 4120 GCCCGTCACGCTTATGAAAAAGCGATAAATTAAGATAAAATCAAGCCATGATTAAAT 4179
QY 110 AspTyArgLysPheTyArgLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrmec 129
Db 110 AspTyArgLysPheTyArgLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrmec 129
```

Db 4180 AATTATGCTGCTTTTATGCAACAGCTGATTGAACAAGCTTCAACAATTTCT 4239
Qy 130 LysProserArgAlaIleGlyTyrGluGlyArgValAlaIleGluAsnMetAlaTyr 149
Db 4240 CAACCATTAATATGCCCATTAATATCAACAGCCGACTACTTAATGAATAATTTAGTTGG 4239
Qy 150 -----IleTyrTyrHisGlnTyrGluAlaAlaLysSerPro 161
Db 4300 TGTGCGCTTTTGGCAAGACACTGTCGTTATCAACA-----AATTGGCAACATTTA 4353
Qy 162 ThrLysAspAspTyrAsnAsnAlaLys 170
Db 4354 ACCGAGTTTATCCCATATGAGCAAAA 4380
RESULT 11
AE000762 13896 bp DNA linear BCT 25-MAR-1998
LOCUS Aquifex aeolicus section 94 of 109 of the complete genome.
ACCESSION AE000762 AE000657
VERSION AE000762.1 GI:2984163
KEYWORDS
SOURCE Aquifex aeolicus.
ORGANISM Aquifex aeolicus.
REFERENCE 1 (bases 1 to 13896)
AUTHORS Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L.,
Graham, D.E., Overbeek, R., Sneed, M.A., Keller, M., Ausley, M.,
Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J., and Swanson, R.V.
The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.
NATURE 392 (6674), 353-358 (1998)
JOURNAL MEDLINE
PUBMED 9819666
9537320
TITLE 2 (bases 1 to 13896)
JOURNAL Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L.,
Graham, D.E., Overbeek, R., Sneed, M.A., Keller, M., Ausley, M.,
Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J., and Swanson, R.V.
Direct Submission
Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego,
CA 92121
COMMENT Putative indicates no similarity to known proteins
Hypothetical indicates similarity to a protein of unknown function.
FEATURES
source
1..13896
/organism="Aquifex aeolicus"
/strain="VF5"
/db_xref="taxon:63363"
complement(10..1116).
/gene="aq_1891"
complement(10..1116)
/gene="aq_1891"
/codon_start=1
/transl_table=1
/product="putative protein"
/protein_id="AAC07705.1"
/db_xref="GI:2984172"
/translation="MISKIKELIQVCEIETWKRLSTGKNGKRAKRTIKYLRKKG
KYVEDFSVETKVPASRIKCEGEFPAPFPIGSIWGTSGEVLKLNLEEKDKG
VVALPVGRRDSEKAPFLDKKASGMVTFLELDVFPSTGIDVYFVAVNTREIVK
IEKGVILIDIREKEIKGKLVYFEGRPGLLVAAHYDKPFFVGAIDNLSVALLL
VMSRELAFQSEIPERIRILPTDCEELGEGASHALSHKNTFYVINDISIGMKPAVL
VEDASGNGRLINELFKLMDKMDYDIPVSKGMSHVFKEKAGQTLFLSDPFT
LHRTQLDLYFAIDMEVAVKWEIVTSYFVRRLHRL"
complement(1106..2041)
/gene="11ve"
/note="aq_1893"
complement(1106..2041)
/gene="11ve"
/codon_start=1
/transl_table=1
/product="branched-chain amino acid aminotransferase"
/protein_id="AAC07697.1"
/db_xref="GI:2984164"
gene
CDS
1106..2041
/codon_start=1
/transl_table=1
/product="putative protein"
/protein_id="AAC07700.1"
/db_xref="GI:2984175"
/translation="MKRVLLLSLILLLSCFPYVORHMKVYDYLGTAAFAARNYSEA
IANFHALRANPDEPRINWALGLAWEAKYKKAESFKALSINPYSARRNLGL
VYLLGRYELAKYLOEANDVEYKHEAFYAKYEAQDLKINRYRYEKAVALP
NFOADLELAQAYENCKYERAKYIKSLINFPNPLKAYEYKQDGERAE
IIVELIKENLTNOREKRELTIVLLAQORLLIIPRHKPIKKEKKKKRYANOL
GASTERADKLVQELSKGLRDLRLPTDGVYKVIYGRPEEARAKEVKKGI
YGVVEIK"
complement(3667..4542)
/gene="foid"
/note="aq_1898"
complement(3667..4542)
/gene="foid"
/codon_start=1
/transl_table=1
/product="methyleneretrahydrofolate dehydrogenase"
/protein_id="AAC07700.1"
/db_xref="GI:2984167"
/translation="MALIDGKSLKIREIKVEYENFTSKGFRPALAVIYVNDP
ASRIYVNNKRAKCEKVGIKSLFYHLPDQVSEKLLGLIYELNNNEVDGLVLPYK
HIDQTVILISPEKVDGFPENMGKLVQIDGFPCTPLGIDILIKRYGIDVYK
DVTIVGAFIVGRPLSLMLMRNATVSVCHIRKDYKFKTEADILISATGVHLIKE
DMIKGAVVVDGISRNGIKIVGDVPERVKEKASAITPVGVGWPTVTALLNLTK
SYRKRFHLISTNP"
complement(4526..5494)
/gene="dnt"
/note="aq_1899"
complement(4526..5494)
/gene="dnt"
/codon_start=1
/transl_table=1
/product="dolichol-phosphate mannosyltransferase"
/protein_id="AAC07699.1"
/db_xref="GI:2984166"
/translation="MISVIVPAYNEENIPVLVEKLLKVLDSLGEVYELIIVDDGSTD
RTPEILKEILANKKQKAVIRFRNYQTALIVAGFEAKGDVITMDADLONDPEIDP
KLIEKMEGDIVSGWKDKRDPPLSKLSKLANMTISVYGVVELHYDCTTAKAYA
DIAMRYLVGDMMHFLPALKRGAKITELPVHHPPLVGRSKYGLDRTIVLIDITL
VKPLNEYITRPLVYFGIGIFLFTLIGLGLVTLVGLFGENIGNRPLILGLVLL
SGIOLISTGIILVIRTVYESGKERYIIIEKINFGVDKDSNT"
5522..6028
/codon_start=1
/transl_table=1
/product="putative protein"
/protein_id="AAC07710.1"

```

/db_xref="GI:2984177"
/translation="MKRIILALFFFLPAFQSYLSPEBEKQIIRIDIAIKATLKTMEQ
TDRFPDRLINELREDNKRFEQVDKREQDREQINNELRLIQMVGIFGO
IALVAIVIGFAMWDRITIRKSKETFEEMKELRPFKLLNALREKAKTDEIA
ILKKYGLL"
6041..6448
/codon_start=1
/translation=1
/product="putative protein"
/protein_id="AAC07709.1"
/db_xref="GI:2984176"
/translation="MKLIDTTLVLDLFSGEKEKVTIEOFPELSQKGLFPEEVI
TELVLFLHGYKMEREDIEVEFTILNDELFNVELKPIREALIKLSKQGTLDCL
SVAKKMGKIEVVSFGRFRFKLGFTKNPYEES"
6462..7910
/gene="gcap1"
/ncore="aq_1903"
6462..7910
/gene="gcap1"
/codon_start=1
/translation=1
/product="glycine dehydrogenase (decarboxylating)"
/protein_id="AAC07701.1"
/db_xref="GI:2984168"
/translation="MELIFKSKKRGKGYLPELDVEEVNIKEYLDEEYLRELDPE
VSELDVRYHTNLSHLYAVDTWVPLGSCCTMKNPINELVNKEFLVHPLTEE
YIOPKLVEYELKELKELGFAVSLQPAAGAGELGLLHAHYHODGNKEKVV
LIPDSAGCTNPASAIQGPDIKVSQSKDELDPEDIKLDERVALMTINNTIGI
FERKIEIAEELHRLDALYMDGNFALYGRFKPQWGVDFVHFKTFTSPHGGG
GPAGPVGVSERLKYLPVPOLEYDGKYLNNWIKESVGKILAFHGLVWLALAY
ILYQGDILKVSSEYAVLNARVLYLKLKGVKDPYSPCMHEFLVSLTNLTLYGVAS
DVAKRIDYGFVATWYFPLVREALMIETETENPDTLKKFALLIKIYKEAKERBE
ILKAPRTVRIRIKAEANRNILKFKDIKE"
7938..9539
/gene="sera"
/ncore="aq_1905"
7938..9539
/gene="sera"
/codon_start=1
/translation=1
/product="D-3-phosphoglycerate dehydrogenase"
/protein_id="AAC07698.1"
/db_xref="GI:2984165"
/translation="MYKVLITDPIAPEGIELLOKDEVEYVNEPDISYEELIIRKQF
DAIITTSRTPTVTELLERAEKLVKVGAGVGVNDIEBATKRLIVVNTPGATIGA
TELTMGMMLTIMGKHAHESMLNYKDKKPMGEELYGIIGIGGNGSOVAIRA
KAFGMKMAVDYITPREKAEKLVKVDNLHMLREIDVITIAPLTHETKMTDEKE
FELMKGVYIVNCARGIINEKALIKMESGKIKGVALDYASKEPPEPIDELKRLA
DKYNISLSPIGANTYESQNVAVITVAQVLKALKQGTVAVANAPPEPIDELKRLA
YLDLAELKGFVQMTBEGIREVHIEVRGIAEFQPISSAVKGLIEEVDPPVNI
NAPYVAKDRGIKVBELESSETPDKHYIKVYVADGKEXVAGTVEGOIPRTIEDR
KYVDIEPEGILVFNKDVGVIGKISIIIGEANINAGRLGRBKKGIAIGILND
EPASSEYLSIKKIPETLFFVXHLV"
9558..10109
/gene="umps"
/ncore="aq_1907"
9558..10109
/gene="umps"
/codon_start=1
/translation=1
/product="uridine 5-monophosphate synthase"
/protein_id="AAC07702.1"
/db_xref="GI:2984169"
/translation="MLMLREKLAARLIKRSKLVADBPVFKSSGKSRVYVLUKQIT
FDEGDLTGKAMVELKERNPDACGLTLAGDPIAIAIFVSLMDSNPILKPVVKE
PKHGKMRQIEGLNPERVALEDVTTGSALAKAVKACREGLVETGVFAVDSBE
GGRNIEKEGIPLYSIFKSELN"
complement(10106..11218)
/gene="hfix"
/ncore="aq_1908"

```

Alignment Scores:
Pred. No.:

0.0609 Length: 13896

```

Score: 122.00
Percent Similarity: 44.13%
Best Local Similarity: 24.02%
Query Match: 11.70%
DB: 1
Gaps: 8
US-10-069-544-2 (1-204) x AE000762 (1-13896)
QY 30 AsnProGlnLeuAlaGlnIleAryThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
||||| : : : : :
2930 ACCCCAGACTACTCGGAGACGAGAAAGATTGATTTCTTACTACCAACTGGAGAG 2989
| : : : :
QY 50 LeuAspGlnAlaLysGlnLeuAspAlaAlaLeuSerAlaAsp-----ArgGlnPhe 67
: : : : :
2990 TACGAAGAAGACCTGAAGTACCTTCAGGAAGCGCAACGACAGTACTGAAAGAA 3049
| : : : :
QY 68 AlaProAlaTyrrArgThrGlnLeuAlaLysValTyrrGlnAlaSerGlnAspAlaThrHisGln 87
| : : : :
3050 CACGAAGCCTTTATTTATTTAGTGGAGGATACGAGCAAGCAGCAACCTTAAAAATTAC 3109
| : : : :
QY 88 ThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSer 107
| : : : :
3110 GTC-----AGGTATTTAGAAAAGCGGTAGCCTTACCACTTGTTCAGGCT 3160
| : : : :
QY 108 TyrMetAspTyrrGlyPheTyrrLeuValGlnMetGlyAspLeuSerGlyAla---LeuIle 126
| : : : :
3161 CAGCTGGAACCTTCACAGGCTTACGAGATTTAGAAAGTACGAGAGGAGAAAGATT 3220
| : : : :
QY 127 Tyr-----PheAspLysProSerArgAlaIleGlyTyrrGlnGly 139
| : : : :
3221 TACAAAGCTTCTTTTGAACGGCTTAAATAGCCA-----TTTCTCAAGTCAAAA--- 3271
| : : : :
QY 140 ArgValValAlaIleGlnLysMetAlaTyrrIleTyrrHisGlnTyrrGlnAlaLys 159
| : : : :
3272 -----CTCGCGAAGTTACTAT----- 3289
| : : : :
QY 160 SerProThrLysAspAspTyrrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeu----- 177
| : : : :
3290 -----AGAAAAGAGATTACGAAAGCGGAGGAGAAATTATTAAGAGGCTCCTTCAAG 3343
| : : : :
QY 178 -----IleSerGlyThrGlnHisAspGlnIleLysLysSerTyrrAspLysLeu 194
| : : : :
3344 GAAACCTCACCAACGACGAGGAGAAAGTAAAGAACTCTCTCAAAAAGTCTC 3400
| : : : :
RESULT 12
AE000777/c 15059 bp DNA linear BCT 25-MAR-1998
LOCUS AE000777
DEFINITION Aquifex aeolicus section 109 of 109 of the complete genome.
ACCESSION AE000777
VERSION AE000777.1
KEYWORDS GI:2984377
SOURCE
ORGANISM Aquifex aeolicus.
Bacteria; Aquificae; Aquificae; Aquificae; Aquificae; Aquifex.
REFERENCE
AUTHORS Decker, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L.,
Graham, D. E., Overbeek, R., Sneed, M. A., Keller, M., Augay, M.,
Huber, R., Feldman, R. A., Short, D. M., Olsen, G. J. and Swanson, R. V.
The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus
JOURNAL Nature 392 (6674), 353-358 (1998)
MEDLINE 98196666
PUBMED 9537320
REFERENCE
2 (bases 1 to 15059)
Decker, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L.,
Graham, D. E., Overbeek, R., Sneed, M. A., Keller, M., Augay, M.,
Huber, R., Feldman, R. A., Short, D. M., Olsen, G. J. and Swanson, R. V.
Direct Submission
JOURNAL Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego,
CA 92121
COMMENT
Putative indicates no similarity to known proteins
Hypothetical indicates similarity to a protein of unknown function.
FEATURES
Location/Qualifiers

```

```
source      1. .15059
            /organism="Aquifex aeolicus"
            /strain="VF5"
            /db_xref="taxon:63363"
            complement(73. .1365)
            /gene="sqr"
            /note="aq_2186"
            complement(73. .1365)
            /gene="sqr"
            /codon_start=1
            /transl_table=1
            /product="sulfide-quinone reductase"
            /protein_id="AAC07903.1"
            /db_xref="GI:2984385"
            /translacion="MACHVVVIGGCGGATATVNLRLNMPDLKTLISDRPYGFTPA
            FPHLMGMRKREDISVPLAPLPKFNIEFINEKASIDPDANTVTQSGKIEVDYLV
            IATGPKLVFGAEGEENSTISCTAEHLAETQKQELYANGPVVIGAIIGVSCFGPA
            YEPALMAYELKKRGIRYKVMFTISEPYLHFGVGIGASKRLVEDLPAERNIDWI
            ANVAVKAIEDPKVIYEDLNTHTEVPKFTMPMSFGGPEVVASGDKVANPANKVI
            VNRCFQNPYKNITFGVGVVTRPIPIEKTPIPTGPKTGMMEQMAVAHNIYDIN
            NPKRIAPRLSATICIADFGEDAGFPFADVPVIEPRKVIITKGMKAHYKTAPEKYLEMK
            VRNGNIAPSEEEKVLEIPLKVHPIELCKDEGAGSRS"
            1544. .2320
            /gene="coxC"
            /note="aq_2188"
            1544. .2320
            /gene="coxC"
            /codon_start=1
            /transl_table=1
            /product="cytochrome c oxidase subunit III"
            /protein_id="AAC07902.1"
            /db_xref="GI:2984384"
            /translacion="MAHETMHEEVSVMPVIGMPLFAPIALTAYFANOMHAGLV
            LGGSLILIFGAAGMANEHFTKDRREGALGTYFPPIIEAVIFGTMFAAYTART
            HAHMANVVEGSIAMAGLTLILMASSYITMAKATILEEGNNGALAKVPEFTPL
            GLFVILHVNEMKLMHGGFTLSNMGTGFYALTGHTSHVLGLIMQYCMWLAT
            GKMQLRPTVLRATSAVWHFVDFMWLLVAGSVYIIGLVG"
            2322. .3119
            /gene="aq_2189"
            2322. .3119
            /gene="aq_2189"
            /codon_start=1
            /transl_table=1
            /product="putative protein"
            /protein_id="AAC07905.1"
            /db_xref="GI:2984387"
            /translacion="MRILLSEFLITLVFGGAVGSPSKAPEDNDMFNPELIKINEQY
            LGNFVNVVEVLEBDGTDLTYLKEKPTILLSYTCGEGCPARVNDLKKLIDETDL
            KDRPMVNLNSFDRDLOTLKEPVKLGHPPTKMWGIIIDKENIEKLINSVGKFPF
            VERKTFVHNVYIFLSPEGRTIYLVNGVRSYDVRIALAESTQNKITVNSYVDLAY
            IACTYDPSRSRYINPTLFGIGFSALGLVLSGLMNLKKEV"
            3122. .3832
            /gene="coxB"
            /note="aq_2190"
            3122. .3832
            /gene="coxB"
            /codon_start=1
            /transl_table=1
            /product="cytochrome c oxidase subunit II"
            /protein_id="AAC07901.1"
            /db_xref="GI:2984383"
            /translacion="MEOVIALPYKMAEAVTWMLIVANVITYLIAIGIYFYKRYK
            FGERIGIDHDEGHHGDEALTTIPLTLVIFLATYSIFPKQRTVPDGAATIKVAF
            MMGQFDLIDENGKVIKTIVVGYHANYNLOEFEPVIPAKSPIRALLTSDVHSFP
            VLPARITEDMWPGRTTYLMQINKRGEVWFCEBYCGTSHMFTKLKVVPPEOFTAM
            LYKNEVAIKSNKKEVAR"
            3994. .5451
            /gene="cox1"
            /note="aq_2191"
            3994. .5451
            /gene="cox1"
            /codon_start=1
            /transl_table=1

gene
            /product="cytochrome c oxidase subunit I"
            /protein_id="AAC07899.1"
            /db_xref="GI:2984381"
            /translacion="MFGQVDFKTYNYILLTHGVGMLEFWMAIAHIGFGNPLPLM
            IGAKDVAFPRLNALSWMAFPAANVLEPLMLIPGNHMTMTWTGPPPSLNNDAGVVALY
            VEVMLLIGASSIATAVNPLVITYIMKAPGTEFKTLFPHTLIAANVLDLVGVSILAG
            AVTMLPLDKYLGTFNFPNPAKGDPILYQNIFMFYSHPVVQVLPVGFGESEIIIPVA
            RRLPFGTSMVFVAWMTAVSFFMLIHMHNFVSGVPKTVVLSYTLTLLIAPVGIKIF
            NMULTIYRGAIHRSQMLYALGVFPELIGLGLDGLGVNLAIDIGVSDSLFVYGPHY
            VIGMAITTLIGLAVIHMFPLATGKIPEMLGKSLFPLIRAGANIPIFLQMVAGMLNP
            RRYDDYPIPEWTLTAVIQTIGALLGVGVVIFTYGIIGLNGSERASDNPMKPSLE
            WKVGVPARAHGQDHPREVDPWHPYRYKEELIK"
            6002. .6259
            /gene="coxB2"
            /note="aq_2191a"
            6002. .6259
            /gene="coxB2"
            /codon_start=1
            /transl_table=1
            /product="cytochrome c oxidase subunit II"
            /protein_id="AAC07910.1"
            /db_xref="GI:2984392"
            /translacion="MMYFDNKGATEIKIPVGSVDIFTTSQVHGVHIGTVNMV
            AIPGVGYMRIKFEKPGVYHVVCHEFCGVGHAMQCKIIVE"
            6265. .8043
            /gene="coxA2"
            /note="aq_2192"
            6265. .8043
            /gene="coxA2"
            /codon_start=1
            /transl_table=1
            /product="cytochrome c oxidase subunit I"
            /protein_id="AAC07900.1"
            /db_xref="GI:2984382"
            /translacion="MMWQVNSAIKFIILTEILFPTLLVFGIYHGMVQFVRSGLIK
            ASBFLGIDVYOGTLHGVINIVYTTIFIVGFSNAVAASLKKPREEKQWALAGMV
            IGTLMAMMFTGRATVLTFFYPLIAHMTFYGAVALVLSLVPEFPFMIPSAIGMK
            RENDOKPLAVGTFVNFILMTIMVPAVAILFOLPLDSGLVDEINILARTLEW
            FEGHPVYFWMLPARYALTYTLPKIYSEGGKLYSDPAALATPLTIFSLPGVLIHOF
            TDPGITNTMKLHALFTFGVALPSMTATVATSLSYKAEHPELKNKSKFYMTLEP
            FMRLEGNKMFSYFFAGVLPFGLGTVGNASYNVLVHNATAVYVGHHTVGGSLV
            LTVFALSLYMSKLGSEVKKLGLAVLAPYEMWQMFMSYAMVGVGVGPPRRTN
            AGLTYNPDSPEYRPMETGAOLAAGVLAIGFAFYASLIATATLAPKVRSTLEF
            PLDAVYHDAPIRLNNTKMTVVAIILVLSYIPDYDSVRSVFFKSPAYNKKFMP
            LKQLQAEKKEKKEKSKABGGITQK"
            8040. .8738
            /gene="aq_2194"
            8040. .8738
            /gene="aq_2194"
            /codon_start=1
            /transl_table=1
            /product="putative protein"
            /protein_id="AAC07906.1"
            /db_xref="GI:2984388"
            /translacion="MKLRSFPLITFLAQTGSTGIPPEAKTLGTYVPGDITLVDS
            YGNERPOLNKKGKPIILSPIYTHGRACAPLTIKLGTPKDGKDEIVTFEEDP
            KOTLEDIKRPQKEYGIDGKGMKYAKATSEDLKLLDAIDPMTGNDRIHNNVVV
            LSPELQIKORYTGVNNTYLFVVALRLARGETALPBGFSYSLFTIGMVGIVGTSYIM
            YLNLRILOKRYKKA"
            complement(8714. .9484)
            /gene="baca"
            /note="aq_2195"
            complement(8714. .9484)
            /gene="baca"
            /codon_start=1
            /transl_table=1
            /product="undecaprenol kinase"
            /protein_id="AAC07896.1"
            /db_xref="GI:2984378"
            /translacion="MTEQDAVVIGIVEISEPLPISTGHLITTAHILGIKHTDPVKS
            FEISIQGSLAVAVVYFNRLIRDYIWKRIIAAFIPGTGIGLLKILKGLIGLNDL
            VVVVSLIGGIIILIPADTYCEKPCYGLDRLRAFMIGVQSIJAVIPGVSRSGST
            IIGGMGLNKRKVAEFSLAIPTMFAATYDILKSGSGFNAGQENNIIIGFTSFI
```

gene
complement (9488..10243)
/gene="a9_2196"
complement (9488..10243)
/gene="a9_2196"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC07908.1"
/db_xref="GI:29843390"
/translation="MAIVLTDRGTGDFGAVKGVITLSINPSVQIVDSHEVDSTNV
LEGALLLKAHYSFPEKSYVGVADPQVSGERGIIVTKETFPVDPDGLFLVLKE
AKDPVEYIKNEKFTLKNNTPHGRDVFAPVAIYSKGVPEEIGRIEIREKLFEP
EPKREKDFEGEIIYFDKFGNALITNVPCKGKVAFAERREDLKAVPYFLAGEKGLNAT
CGSFGMEIRPVDPNAREKNLKEGKIKFFII"
complement (10243..10731)
/gene="a9_2197"
complement (10243..10731)
/gene="a9_2197"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC07909.1"
/db_xref="GI:29843391"
/translation="MLAMEGLKEQAMKLFEEAYKYHMLGDLDAIEIYKRSIDVPTA
EAWTFLGMAYSMRNGEAIACKRAIEIDPDPGNIDGSLYIEGLKDEAIEMLE
KAKARARYEPRHYPYINLAKVMLOGLVLALEIEBAIKIKRDPYKSHLRHQLIGM

Alignment Scores:

Pred. No.: 0.076 Length: 15059
Score: 121.50 Matches: 39
Percent Similarity: 47.90% Conservative: 18
Best Local Similarity: 32.77% Mismatches: 40
Query Match: 11.65% Indels: 22
DB: 1 Gaps: 4

US-10-069-544-2 (1-204) x AE000777 (1-15059)

QY 70 AlATyArGhrhLeuAlaValaTyrgAlaLaseGluApla----- 84
Db 10645 AGCCATGAACTTATAGAAAGCTATAGAGCTTTACCCACAGCCGAGCTGAGACTTT 10586
QY 85 -----ThrHISGIntHrLalaglnArg-LeuPhegu-----Ly_96
Db 10585 CCTCGCTGGCTTACGACATGAGAGAAATTACGAAGGGCTATAGAGCTTGCAAAAG 10526
QY 96 aAlaIleGluLeuAnPrOlysaPheGlnSerTyMetAspTyrGlyPheTyrLeuVa 116
Db 10525 GGCTATGAAATAGACCTTGAATCCCTTACACAGACATAGTTCTTACCTGAT 10466
QY 116 lGInMeGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleG1 136
Db 10465 AGAGCTCGAAACTGACAGAGCCATGAGCTGGGAAAGCTTAAAGGGCTAGAG 10406
QY 136 yTyrgLugLyArGValaValaIleGluAaMeAlaTyrlleTyTyHISGIn----- 154
Db 10405 GTACGAACCTAGGACATCTTACATTAATCTCGCAAGGTTACATGTTACAGTTAA 10346
QY 155 -----TyrgIuaAlaLaseSer-----ProThrlsAspAspTyr 166
Db 10345 ACTCTACGAAGCTTAAAGAAATAGAAAGACTATTAAGATTAAAGCCGATTAC 10291

RESULT 13
AP003190 301150 bp DNA linear BCT 10-JUL-2002
LOCUS Clostridium perfringens str. 13 complete genome, section 6/10.
DEFINITION AP003190 BA000016
ACCESSION
VERSION AP003190.2 GI:18146728
KEYWORDS
SOURCE Clostridium perfringens str. 13 (strain:13) DNA.
ORGANISM Clostridium perfringens str. 13
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.

REFERENCE
AUTHORS Shimizu,T., Ohtani,K., Hirakawa,H., Onshima,K., Yamashita,A.,
Shiba,T., Ogasawara,N., Hariori,M., Kihara,S. and Hayashi,H.
TITLE Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
MEDLINE 21664373
PUBMED 11792842
REFERENCE 2 (bases 1 to 301150)
AUTHORS Shimizu,T.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical
Sciences, University of Tsukuba, Department of Microbiology; 1-1-1
Tennohda, Tsukuba, Ibaraki 305-8575, Japan
(E-mail:tsheimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354,
Fax:81-298-53-3354)
COMMENT On Jan 14, 2002 this sequence version replaced gi:18144942.
FEATURES
source
gene
CDS

gene
complement (3..5066)
/gene="CPE1281"
complement (3..5066)
/note="1687 aa, partially similar to pif_736472 probable
secreted alpha-galactosidase from Streptomyces coelicolor
(680 aa); 33.6% identity in 140 aa overlap. Putative
N-terminal signal sequence was found by PSORT"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80987.1"
/db_xref="GI:18144943"
/translation="MNKRKIAAIIATMTNLSATTIDVLAQELNKNKSVESVSHD
ESHQARVSKFEDLYNSDKLDAYNOEFOINSNKSITNNGKXNSSTIDRAIDNLETH
WETGKQNSDFNNEVYVTFNEITNIDRIYVSAROSARQKKEEIEIADQDEDD
NIVSSGCTESTRDVETKRNPTDPRKLRKFKKADQWMASSAEPMFKEDLNKEF
NLGFTDSNNKSVSEFNTLEKLNAPENELKDHPMYDLYKEGLNNAAILTESENPK
ATLGOITYNLNDYNNQYRMYPTNIFKSIKNNRSHYAQNIKAIIDNDVNTWETGLN
SSSFNNEVEFNDVLTLDRIYVSGROSOLKFAEBEYVYASRTSGDPTKYLAVATGS
BATKGLVEAKEPEPTPEKRFKPKKSQNSATINELMEFKNPBDEVSISPKLFTDGTMS
ELSEFNSLEKINAPKERAKNHPIYNDPVDIADLSLISNPKEDVLEMGDSIS
EAKRKVYANFQWQITGTSARAGDKTYVVDVDAEGDPTTLTKOSLTQHGATSTQD
KQKQETITPEINYESNGIPKQVIOGGDLFTNYSKDSQKRAKVAIEGASKYPTIL
GKSDENWKELEAVYEKIKAEPTTPNIPFAVSNNKSLFEVQATVADLYKKNKTRPK
YVAEQMDQYIADAMGFWGPDNSKDVSNDENPIMPVKULSGAPNAGNGVIGIRG
NODAILAANKGQVHLEIGHNFDTGRTIVETVNNMMPLEFESKYKTKRTITPONIME
NNTYPRVGLDVSNNELYNKADSTHQAQLPMLQLYLVNTPEKREPRERENDPKNK
NNEBIVKSNVVAASDAMEIDLTFFRKHGRVDDKVEDLAKPKRDKKIYINDLAM
NKGGDFTNNAVYVSTSGSNINIKLSFVDDENKNIIGYEIRRGKVGFTNSNSF
VDTKSLDEDDGVYVTPYDKLNTLNPISVNALOPTLSVPVYITTLAGEFNEEYIV
ADQIKNSLSSEVYKSVNVSKEVEYVLSLEBQKNEBTKTSKYSNASKYMS
DLPPOKSNMGKTVRKDSISGGVIGLTGDFGVQVDYKGLSHNAEYVYDLEGKYD
YRESYGVDRKAMSSRPASSVITKVLVDGEKERSGMSRTTQKYKVDVYKAKKEKL
IVNDAGDSDSHASFGAKLATLSKPIIKGNSLAINDEKEDLMLKGLTIADIEGN
ITSKIQIKSSDVEBEGSKGFKVYVSYTDSDGLTSGSRTIAVDKETOJSLDMWKAT
IGSGVYKDRVAESGNQIRLNLNEDNSVQTFAGIEGSRYSSEIYVNSGDIYFDTWGID
RHVADKVSYSVEFKYVVDGELKAETDVMIIDPPKRLVVDVNRSKIKLVVDVADNG
TWHDHMDAKKRPNLAEVDTTELKALAEAKGLDINNRESEFALKNAISKEEALL
SKDKRTINSALEIKAMDSLYKVDLNNATINIPDKYKLSIQNLKQTKGDTGMYVS
LTTTLISGVEDLSLENAKNIETLNDNDNEVDLRLSLKSLKNTLNTNABOQFAABEL
KPSNGKVIDGSKYVNRKGVAKTIRVVDKNGNTLEODAKOEFITNTXDLSSGLYGV
HVLFBDBGSGVMFYLFNV"
/gene="CPE1282"
complement (5515..6903)
/note="462 aa, similar to pif_2516401T transmembrane
protein from Bacillus cereus (441 aa); 25.5% identity in
377 aa overlap. 13 putative transmembrane regions were

gene
CDS

```
found by PSORT."
/codon_start=1
/transl_table=11
/product="probable transmembrane protein"
/protein_id="BAB80988.1"
/db_xref="GI:18144944"
/translation="MDISLNTLENVEEYKKAQPAVALMIFALVDNKRISPOL
AKANISATVAVSQVTLGAVIGAVYASLSDISVRLVLVGLICVGLIGT
FOKSFLLIVISRIIQSAGLSAETLVYIFAKYKSEHKYLGSTSYAISQVIGT
ITGVYISTYLSMQLFLPLTLVLPLFLYIPREKKNAGNDYIGFLLVATVAATL
MIFVSGOMGLIGFIVALLFLYISKNENAFISKEFTNRNFISVLVAVYVSQ
LAFIFMPELVEISYNYRDLQISLIVSVYLAATVAGLSGVSKKISKOCIMAIY
GLIYSLIIGALFVKTSVYIFPMWILPSSSALVAMIDILTRTIPREKGTAGY
NFCNLNATISGITVVALFMETKYLGNFTGLSDPLVALQISNVLIVAVLSVALVY
WTLIGRREKKOI"
/7322..8071
/gene="CPE1283"
/7322..8071
/cds
/gene="CPE1283"
/7322..8071
/length="249 aa, similar to gpu:AP001515.128 transcriptional
regulator (GntR family) from Bacillus halodurans (242 aa);
24.9% identity in 229 aa overlap
GntR family"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="BAB80989.1"
/db_xref="GI:18144945"
/translation="MKLTTKDTKTSKLLYTSVYDKYKMIIDGEYPLNSKLPSEPEL
AKSPGVSMTRQALALDDGLVKNIGKGNFIISPEKARDFLELIGNPIYKCHT
EKIDIEFVPELIEDSDYTKOVLKEDSTVAACRYKSKKVALAVATFISVETVE
IKIDLNSESLDLPEHKSYSIESSTIEIYHSSAGNSLNKRYKLVGNQCDVLETL
YANDISPELTFNKYIIPSOYSTFKIKARKE"
/complement(8189..9370)
/gene="CPE1284"
/complement(8189..9370)
/length="93 aa, similar to gp:P45912.25 pyrimidine
nucleoside transport protein from Bacillus subtilis (93
aa); 71.7% identity in 92 aa overlap. Putative N-terminal
signal sequence and 7 putative transmembrane regions were
found by PSORT."
/codon_start=1
/transl_table=11
/product="pyrimidine nucleoside transporter"
/protein_id="BAB80990.1"
/db_xref="GI:18144946"
/translation="MKYIIGVGLIVIALAMIGSDKKIKYKVPFMLILQFVLG
ILNTOGVNYLVGSIANGFOVLDDGAGVAFVFGIANNDOFSEFISLLPIFISA
LIGILOVYKTIIPVYIKYGLSKINGMKLESYNAVASALIGSEVPIISVKEGLIL
PKHRLYTICASAMSTVSMSTVGSYVWLQIPRYVYVATYLANLFGGFTIASVNPYSVA
BEDILVDEKQKQTFEVLGEYILDGFEVAVIYVAMLVGFVALIIVAIMIPRSIFGMS
FOEILGYFSPFAPLGMVPLNESIQASVMAKYSNFEVAMNLLSGTILNLSRSGIS
IVSVFLISFANFSSIGISGAVKGLNQGVNVARFGLKLVGATLVSMLATIVGLI
I"
/complement(9831..10928)
/gene="CPE1285"
/complement(9831..10928)
/length="365 aa, no significant homology."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80991.1"
/db_xref="GI:18144947"
/translation="MENTWEKKWVEDIYELKEELKRRNNLFAYTEESFNEKLENL
MNPENLDEBEIKVEISRYVASLKOHTSLIPAKPFIPLKTYFNDGVIYINTCKGVE
NLPRKVALIGDMKTEVLELSNLSISENBYFPQSQMOQIAEVYIITLIDTMD
KITITLDGEYEVLTCSFEDLTYNKRIPWAKNDSEMLMFOIGSGCLYIKNSCRE
QGESEIERKIEINILCIKENIEKTVLDRNNLGDSTLFTPLIDYLNKSKINKEN
LKVIGRETFSSALNATYFNSTNAKTIIGSPSGPCVGEIIRITLIPNSKLVITYS
TRFYKLEDSVMALYPREVLIESIEDYINK"
/11256..11708
/gene
```

```

CDS
/gene="CPE1286"
/11256..11708
/length="150 aa, similar to gpu:AE004385.4 conserved
hypothetical protein from Vibrio cholerae (170 aa); 43.8%
identity in 144 aa overlap. Putative N-terminal signal
sequence and 1 putative transmembrane region were found by
PSORT."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80992.1"
/db_xref="GI:18144948"
/translation="MDVLYVVALSVGLVGVMAAMAFAVGMVMTGVGNTSYSSGG
MANGLFRLSCSNATGIIYALLIITKSSIPFNEVNSFLTRAPFPMVYIAKPKKPEFI
SGAFCCGFTFGINGMVAIVPALCGNVGLYISQIKIYCHKVFSTKI"
/complement(11999..14614)
/gene="CPE1287"
/complement(11999..14614)
/length="871 aa, partially similar to pir:C72285
hypothetical protein from Thermotoga maritima (strain
MSB8) (718 aa); 22.8% identity in 487 aa overlap. 5
putative transmembrane regions were found by PSORT."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB80993.1"
/db_xref="GI:18144949"
/translation="MILLYNKKRYKOEYXNLIFKGNLTLLISCFPMIGLIFLDNVNF
ILKTYOENKLSVALILFPIILLVGLVEKLNIIKRIINDIDPLAISLSISILII
IYGIILKIDNRYKIEVSLFVITIIIGVIRIVINAMKNEKTNITLKDLYENKI
SLDSHPIFLNEEESYDLLENRNIIIEKYEALVNCNPKRFTIISLGMNGSKTTL
NIVSKKINDNNKDIKIIISFDPWINDQISWFSMDIILKGTGISYISCTKRLVND
IYNIIPSTKYTKGIKDLNFENHDKTEIEKMKMINNYHISNKRIVFIIDNIDRAEK
ENIILFELKLVANFNEFEVYVYILSPDNKDKLIEQLDIDYFISKIVQLPIKIPPL
DEKNEVISTCFKNIIIRLYGEONLEKYNLINSLSKLIDMDPFRKISVSVYHK
NCEMINDIDISIEILFENYKDIYISIKENPKYISPLIDSEYVIVDKYENIDILAKM
GKXFPDKLFENNPFIFKEILSVYFVYKRYNSYSLIDSEYVIVDKYENIDILAKM
RITSARFSLYFSEYKNEFINISKSIDDFNLVNNNSNETTEKILCTLEAYCSMO
ETLIFLQNRIDINDSDISVFTKALEENINIFDSDLVFGLNGLTREIETIISMLLE
LNNPELEIKFELIKDYKIKIFINSIKYWNSEVSKNLEQILFLENLELLVYEV
LNNNIDYKSYVYRGNLRYIEKIVANDKNYIKKYVGLIENNNIIFLFPMSISVGT
TYKYFIRBTVLDMSSKEEVEDELIDRCKSFNEDEKFLVDYKKMKLKDKNEDMNMND
DDGYLMISIKININ"

Alignment Scores:
Pred. No.: 4.49 Length: 301150
Score: 120.50 Matches: 43
Percent Similarity: 47.16% Conservative: 74
Best Local Similarity: 24.43% Mismatches: 40
Query Match: 11.55% Indels: 19
DB: 1 Gaps: 6

US-10-069-544-2 (1-204) x AP003190 (1-301150)
QY 30 AsnProGlnIleuAlaGlnIleArgThrGlnIleAlaIleSerLeuAspMetGlyLeu 49
Db 22958 AATCTTATTTTAAAAATGCTTATCTGGAAGGCAATATTCGCTTTGAAAAGGGTGAT 23017
QY 50 LeuAspGlnAlaLysGlnGlnIleuAspAlaAlaLeuSerAlaAspArgGlnIleuAlaPro 69
Db 23018 TATCTTCTTGCCCAATAATATCTTCAATATTTTGTTCATGTAATAAATAATATGCTTT 23077
QY 70 AlaTyrArgThrIleuAlaLysValTyrGlnAlaSerGlnuAspAlaThrHisGlnThrLys 89
Db 23078 GCGTATATAAATAATAGAGATGTACTTCTTTTAATAGACTTAAAAACAGTCTTAA 23137
QY 90 AlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMet 109
Db 23138 -----ATGATAGAGAAATCTTAAAGAAATGACCAAGAAATGATGCTTTCTTATGCT 23188
QY 110 AspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheasp 129
```


translation="MYTISINFCRVIQDTKKMKMPTEKKPALINIRAKKEID
 LLEFIRKIAFEKLTUATETTESMEGKPEVEFPAIDGPAFTVEFNH
 STEVKGQOILEDIFVKRPGKIGIKAMFHCINIAKERNCGMEWAVIDMNPAREF
 YEVKAGPVGDMHLYRMSADKESALK"

BASE COUNT 3317 a 2075 c 2274 g 2641 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.0843 Length: 10307
 Score: 119.00 Matches: 56
 Percent Similarity: 40.25% Conservative: 41
 Best Local Similarity: 23.24% Mismatches: 72
 Query Match: 11.41% Indels: 72
 Gaps: 8

US-10-069-544-2 (1-204) x AE013241 (1-10307)

QY 12 AAlaMetGlyLeuValLeu-----SerAlaCysGlnSerThrPro 24
 DB 3387 GCAAAAGGCTTGTCTTGCACAAACTTGAATAATGACAGTCTGGAATGCTTGTAT 3446
 QY 25 IleProPcLySAsnAsnProGlnInleuAlaGlnIleArgThrGlnIleAlaIleSerLeu 44
 DB 3447 TCCCTTACAGAGGAAATCCAGAAATGAAAACGCCCTTGAACGAAATGCTTTGTCTT 3506
 QY 45 LeuAspMetGlyLySLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAsp 64
 DB 3507 GCAAAAGTATGGAAAAAGATTTAGCTCTTGAAGCTCTCGAAGATTTCTTA----- 3557
 QY 65 ArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAspAla 84
 DB 3558 -----NAGAATATCCCGCAAAATGAGGCTGCA 3584
 QY 85 ThrHisGln-----ThrLysAlaGlnArgLeu 93
 DB 3585 CTTTACCAATAAGATATCTCTTACGAGCTTTCAGATACGAGAAAGCAAGAAAATC 3644
 QY 94 PheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPhe 113
 DB 3645 TTTTCGAAAGTATTAATACTCGATCCCGAAGACAGGAACTGTGTTTCAAGAAAGTTT 3704
 QY 114 TyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLys----- 130
 DB 3705 GCCCTGTACAGCTTCTCAGGCTCAATGAGCCATTAAGCTTTGAAAGACCTTAAG 3764
 QY 131 -----ProSer-----ArgAlaIle 135
 DB 3765 ATCGATCCCTTATTTTGAAGCCTGGAACGATGTTTTCACCTTATGAACCTCGAG 3824
 QY 136 GlyTyrGlnGlyArgValAlaIleGlnAsnMetAlaTyrIleTyr----- 151
 DB 3825 GTCTACGAAAGAGCCCTTGAAGCTTTGATTCATGCTCAGAAATCTATCCGATGGAAG 3884
 QY 152 -----TyrHisGlnTyrGlnAlaAla 158
 DB 3885 GATATCTGTACAGAGGCACTTGTCTTGAAGAACTGCAGAAACATATCCGAACTGTC 3944
 QY 159 LysSer-----ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnLys 175
 DB 3945 CAATCTTTGCCGTTTACTGAACTGACCCCTGAAAATAAGATGATGCTTCACAGCG 4004
 QY 176 AlaLeuIleSerGlyThr-----GlnHisAspGlnIleLysLysSerTyrAspLysLeu 193
 DB 4005 GGGCTTCTGCTTCCAGAAACAGGGAAGCAAGAGAAAGCCCTGAATGCCCTTGAATAACTG 4064
 QY 194 Leu 194
 DB 4065 CTT 4067

RESULT 15
 U32721 U32721 10507 bp DNA linear BCT 29-MAY-1998
 LOCUS Haemophilus influenzae Rd section 36 of 163 of the complete genome.
 DEFINITION

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

U32721 L42023
 U32721.1 GI:1573334
 Haemophilus influenzae Rd.
 Haemophilus influenzae Rd
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Haemophilus.

1 (bases 1 to 10507)

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,
 Kirtress, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,
 Merrick, J.M., McKenney, K., Sutton, G.G., Fitzhugh, W., Fields, C.A.,
 Gocayne, J.D., Scott, J.D., Shiley, R., Liu, L.I., Glodek, A.,
 Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,
 Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M.,
 Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L.,
 Geoghegan, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M.,
 Smith, H.O. and Venter, J.C.

Whole-genome random sequencing and assembly of Haemophilus
 influenzae Rd

Science 269 (5223), 496-512 (1995)

95350630

7542800

2 (bases 1 to 10507)

Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S.,
 Borodovsky, M., Rudd, K.E. and Koonin, E.V.

Metabolism and evolution of Haemophilus influenzae deduced from a
 whole-genome comparison with Escherichia coli

Curr. Biol. 6 (3), 279-291 (1996)

96398784

8805245

3 (bases 1 to 10507)

White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
 Direct Submission

Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

4 (bases 1 to 10507)

White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
 Direct Submission

Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

The H. influenzae sequence has been updated by R. Fleischmann. New
 database matches have been assigned, product names have been
 improved, and a number of frame shifts have been corrected. We
 gratefully acknowledge the work of Tatusov et. al. We have
 incorporated their annotation into the /notes fields of the
 corresponding H. influenzae genes

5 (bases 1 to 10507)

White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D.,
 Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.

Direct Submission

Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

The whole genome was shifted by 588 nucleotides for a new start
 on Sep 30, 1996 this sequence version replaced gi:1221035.

location/Qualifiers

1..10507

/organism="Haemophilus influenzae Rd"

/db_xref="taxon:71421"

81..620

/gene="HI0366"

/gene="HI0366"

/note="similar to PID:1162960 percent identity: 33.59;
 identified by sequence similarity; putative"

/codon_start=1

/transl_table=11
 /product="Fimbrial biogenesis and twitching motility
 protein, putative"

/protein_id="AAC22024.1"

/db_xref="GI:1573335"

/translation="MYTISINFCRVIQDTKKMKMPTEKKPALINIRAKKEID
 LLEFIRKIAFEKLTUATETTESMEGKPEVEFPAIDGPAFTVEFNH
 STEVKGQOILEDIFVKRPGKIGIKAMFHCINIAKERNCGMEWAVIDMNPAREF
 YEVKAGPVGDMHLYRMSADKESALK"

gene
LEKLRQDGKRAEKNKNSLK"
692. .1603
/gene="H10367"
692. .1603
/gene="H10367"
/note="similar to Sp.:P27434 GB:D21149 GB:U02965 GB:X64451
PID:41541 percent identity: 26.33; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAC22025.1"
/db_xref="GI:1573336"
/translation="MNPTTEOVLSFGEIFRQTRBALNLSIEDVAKEITLRPSILEOLE
KNEFLEKSPALFVKGVRYSVAKFLRLDPSYWEINIVFAETEKNDLGNASTRAWNOY
SHEHNKIGTALIVPMIVIGMTGLMMQSYOQNTQERDLYSVASTENNQPTALIV
TSEENKIVTPEAAVPSQVPEITNNLRLPEIKQNSVSPQPKDKESVSEIQSAVENPI
SPTLIACDQDIYEILTNSSWTSYCDNARHLYLQAKETKQSEILTFNGNRSILVIGAS
NRRITYKGENYELKVDGRVAKFKLSQP"
1613. .2719
/gene="H10368"
1613. .2719
/gene="H10368"
/note="similar to Sp.:P27433 GB:X64451 PID:41542 GB:U00096
PID:1788863 percent identity: 88.15; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="gcgc protein (gcgc)"
/protein_id="AAC22026.1"
/db_xref="GI:1573337"
/translation="MSAFQPTIKRRESTRKIYGVNVPVIGGDAPVAVQSMNTTRTTDVEA
VTAQKLSIEREGADIVRSVPTMDAAEAFKQIKQOVNPLVADIHFDYRILAKLAEYG
VDCILRNPNIGREDVRVAVDCCARDKPIRIGVAGSGLEKDLHGEYRTPALILE
SALRHVELITLDFPOFRVSKASDPLVASVYLAKIKQPHLGLITGAGARBA
VKSANGCLMLAEGIDQLRYSLAADPPEELKVGFDILKRLIRSGEINTACTYCSR
QSFADIGVNALEQRLIEDITIPMDVSIIGCVNNGEALVSDLGATGANKSGYVLDG
EQOKERPNEDIVNLEAKIRAKVADOPKNNRII"
2729. .4000
/gene="H10369"
2729. .4000
/gene="H10369"
/note="similar to GB:M11843 Sp.:P04804 PID:146372 GB:U00096
PID:1788861 percent identity: 66.82; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="histidyl-tRNA synthetase (hiss)"
/protein_id="AAC22027.1"
/db_xref="GI:1573338"
/translation="MAKTIQAIKGNDCAPTESPLMOMIEAQNVTNLSYSGSEVMP
IVESPLPRAIAGEVTDIVSKEMTYPMNDQQLTRPBGTRGCRRAIRHGMITNNEQ
RLMYIGPMRHRPQKGRYRQPHOVGEVFIANPEIDAEILMLTVLMLKALGIDQYV
TLQNSISLELRANYSALVGFLENHODLNSDEKELVNPLRIIDTQNEPILQKVL
DNAKPLLYLDIDESTRHFEOGLSLDAGIQIYEINPKLVKGLDYNVTAEWVLSALG
AGCTVCGGRYDGLVEQGGHATPSIGFAMGLELRLVLLQVENNVPAKSAVIDIYVY
QEGEGLTIAFELAEKVRSELPHLNTMLTCSGNGPKQPKRADSKGATLALVIGSEVQ
NKQVVKHLQSGTDDQLDVLNIIIDYLOTQ"
4017. .4631
/gene="H10370"
4017. .4631
/gene="H10370"
/note="similar to GB:U00096 PID:1788860 percent identity:
36.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAC22028.1"
/db_xref="GI:1573339"
/translation="MAYSIIEEOEINQJLKMWKENGKTIYAFILYIGMGFGRMYQOT
HOAEQIAQSAVDTLINSVQODEQAKKANIIEQVQANSTKRAVAVFALLDEAKKATK
QDFSAENMLNQTOSQDEVLTYSIVALRSLAVQFOGLDNLALSTLNQVKGESFNAR
KAILTGCDIQVANGDGVAAKNSFEQSQGSGOLEQOQMKMKLNLN"

```

gene      complement (4682. .4876)
          /gene="H10371"
          /complement (4682. .4876)
CDS
          /gene="H10371"
          /note="similar to SP:P37096 GB:U00096 PID:1788873 percent
          identity: 73.44; identified by sequence similarity;
          putative"
          /codon_start=1
          /transl_table=1
          /product="conserved hypothetical protein"
          /protein_id="AAC22029.1"
          /db_xref="GI:1573340"
          /translation="MKTKDQQLAEELYDRNPDLDPKTVFTDLHKICLENFDDDP
          NKSNESTLEALILKMWDEFE"
          /complement (4876. .5217)
          /gene="H10372"
          /complement (4876. .5217)
          /gene="H10372"
          /note="similar to GB:M88654 SP:P25528 PID:145915 GB:U00096
          PID:178874 percent identity: 64.55; identified by
          sequence similarity; putative"
          /codon_start=1
          /transl_table=1
          /product="ferredoxin (fdx-1)"
          /protein_id="AAC22030.1"
          /db_xref="GI:1573341"
          /translation="MPKIVLPNEDFCPEGAVDAATGDNLLVFAHAGVEIHHACDG
          SCACCTGVRSQFSDLNSTSDQEBMDLKAWGLEMDSRSLSCQCVVGNEDLVVEIIPK
          YNLNHNAAAH"
          /complement (5257. .7116)
          /gene="H10373"
          /complement (5257. .7116)
          /gene="H10373"
          /note="similar to GB:U05338 SP:P36541 GB:U01827 PID:402675
          PID:454766 percent identity: 69.37; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=1
          /product="heat shock protein (hscA)"
          /protein_id="AAC22031.1"
          /db_xref="GI:1573342"
          /translation="MALQLAIEPQQAAPHQHRILAVGIDIGTNSLVASVRSQSVIL
          NDBRESLVSVVHGVGEERKGLLEAFQESLDPKNTVLSVKRLIGRSLSDVQSRSS
          LPPEFVSENGLEPLIIIPAQGAQPIEVSDDLRLNHLIAEORLGEISGVITVPAYF
          DLRQSTKDAARLAVGLNDRFLNLEPAALAVGDSGGGRIIAVVDLGGCTDLSL
          RLKSGIFEVATGSDTRLAGSDRDHLIADNVITQTKLKPRTANQOIELITLAAQKTT
          LTNKSAVISWQDSVSEISKEQFNELIYPLVKRSLTCRPAALDANVESEVAQVWV
          GGSRRVRYEVOVEFPGKPLTSLIDPKVAAAGAIQADILGNKTSMLDLDPV
          LSLDIEITMGALVEKILIPRNTTIPVARQSEFTFKDGTAMTVHVLGSERLVADCSLS
          GFLRGIIPMAAGAAHIIKVYQVADAGLSVTAMEKSTQVQASIQIKSPYGLDESV
          TARKFSFNNARODLRAELAEORVADRYEISVIYALQADAGLIDLTDFEHHITETVL
          KQMDVYQGSDDRDLIAQGITALDPTAIDQEPARRMNASINKLTKSLDIENP"

Alignment Scores:
Pred. No.:      0.0972      Length:      10507
Score:          118.50      Matches:      39
Percent Similarity: 39.47%      Conservative: 21
Best Local Similarity: 25.6%      Mismatches: 89
Query Match:    11.36%      Indels:      3
DB:             1          Gaps:      1

US-10-069-544-2 (1-204) x U32721 (1-10507)

Cy      1 MetLysIleArGValLySTrPProMetValMeCAlaMeCglyLeuValIleuSerAlaCys 20
      |||||
      |||||
Db      81 ATGAAACAATTAAGTAACAACCTAAGTCGCGGTATTTTCCCTTCATTTTTCGCTGC 140
      |||||
      |||||
Cy      21 GlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArgTnGlnIle 40
      |||||
      |||||
Db      141 GTTTCACAAATTCGATCTTACCTTAATCAATCAAACTGCCCTAAAGCGCGGTGGAATC 200
      |||||
      |||||
      41 AlaIleSerLeuAspMetGlyLyLeuAspGlnAlaLysGlnGlnLeuAspAlaAla 60
      |||||
      |||||

```

Db 201 GCTTGAGCTATCTTCAGCAAAATATCTCACTGCTGCTAAATCATTTAGACAAAGCA 260
Qy 61 LeuSerAlaAspArgGlnPheAlaProAlaTyrThrGlnLeuAlaValTyrGlnAla 80
Db 261 CTTCAGACAGATGAAATATCTATCTCGTCAATTCAGACATTCGACATTATATCAACAA 320
Qy 81 SerGlnAspAlaThrHisGlnThrTyrAlaGlnTyrLeuPheGlnTyrAlaGlnLeu 100
Db 321 CAAGGCAATAGAAAAT-----GCTTTCGTGATGATGAAATATGCGGTAACCTT 371
Qy 101 AsnProTyrAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAsp 120
Db 372 AATCATTAACAAGGCGATGCGATATATTTGGTAGCTTCTCATGTAGTCAAAAGAA 431
Qy 121 LeuSerGlyAlaLeuIleTyrPheAspTyrProSerAlaGlnAlaGlyTyrGlnGlyArg 140
Db 432 TTTCAGACAGCTCAGCAACAATTCGATTAAGCATTAATTCGCCGAAATATTATCATCA 491
Qy 141 ValValAlaIleGlnAsnMetAlaTyrIleTyrIleTyr 152
Db 492 GCAGATACATTTGAAATATATGCTGCTTGTGCTTAT 527

RESULT 16
AF363030 3141 bp mRNA linear ROD 24-APR-2001
LOCUS Mus musculus UDP-N-acetylglucosaminyltransferase mRNA, complete
DEFINITION
ACCESSION AF363030
VERSION AF363030.1 GI:13775065
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Rumberger, J.M., Wu, T., Hering, M.A. and Marshall, S.
TITLE Molecular cloning of the mouse O-GlcNAc transferase
JOURNAL
REFERENCE 2 (bases 1 to 3141)
AUTHORS Rumberger, J.M., Wu, T., Hering, M.A. and Marshall, S.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2001) MCB, Hexos Inc., 22118 20th Ave SE St.
G-142, Bohehl, WA 98021, USA
FEATURES
Source location/Qualifiers
1..3141
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/tissue type="skeletal muscle"
1..3141
/note="OGT, O-GlcNAc transferase"
/codon_start=1
/product="UDP-N-acetylglucosaminyltransferase"
/protein_id="AAK39123.1"
/db_xref="GI:13775065"
/translation="MASSGVNADVSTETPTKMLSRGQIAELAHREYQADPEAAERHC
MOLRQEPDNTGVLLLSIHPOCRRLDRSHFTLAIKONPLABAYSUNYKER
GQLQDAIEHYHARLKRDPIDGYNINLAALVAAGDEGAVQVSAQVNPDIYCR
SDLGLKALGRLEBAKACVYKALETQNFVAVMSNLCGVNAGEIWLAIHNEKAV
TLDPIFLDAYINLGNVLEKARIIDRAVAAYRALSISPNHVVGNLACVYEGGLID
LAIDPFRALIELOHPHPAYCYNLANLALEKESVVAEACQYATRLICPTHSDNNLA
NIKEQGNIEERAVLYRKALEVPEPRAAHNLSAVYQOQKLEALMYEALRIASP
TPDADISMGNTLTKEMDVQALQCYTATQINPAPADHNSLSIHKDSGNTPEALA
SYRTALKLKDPDPAYCNLAQLQIVCDWTYDERMKLVSIIVAQLEKNLPSVHPH
HSMYPLSHGRFRKAIAERHGNICLDKINVLKPPYEHKDKLSGRLARVYSDRG
NHPSTLMQSIIPGMNPDKPEFVFCYALSPDDGTYRKYVMAEAMHFDLSQIPCNKA
ADRIHODGIHILVMNGYTKGARNELFALRPAPIQAMLVGPGTSGALFMQYITDDE
TSPAVARQYSEKLAMPTFRIGDHAMPHLKKAVIDKRSNGHYDNIYVANGID
LKAELDSLDPVKIYAKKCPDGGNDPSSNTLMNPFVMTATIAEVEMIRGOIQT
INGFSISNGLAVTQINNKAATGEVPRITIVTSQYGLPDRALIVCNFQYKIDPS
TLQWNTANLKEVPNSVILWLRFPVAGEBNIOQYQOMGLPNDRIIFSVAPKEHVRA
GQLADVCIDTFLCNGHTTGMDVLMAGTPEMTMGPTGLASRAASQITLGLLEIYAS
RQYEDIAVKLTGLDLEVLKRIKGVKMKRISSPLENTKQYTMELERLYLQWHEHYAAG

BASE COUNT 876 a 705 c 732 g 828 t
ALIGNMENT SCORES
Pred. No.: 0.0261
Score: 117.50
Percent Similarity: 43.09%
Best Local Similarity: 25.53%
Query Match: 11.27%
DB: 10 Gaps: 7
US-10-069-544-2 (1-204) x AF363030 (1-3141)

Qy 10 ValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProProTyrAsn 29
Db 397 TTGTATACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 456
Qy 30 AsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
Db 457 AATCTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 516
Qy 50 LeuAspGlnAlaLysGlnLeuLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
Db 517 TTGGAAGACCAAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 576
Qy 70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAsp-----AlaThrHis 86
Db 577 GCCTGAGATATCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 636
Qy 87 GlnThrTyrAlaGlnArgLeuPheGlnTyrAlaIleGlnLeuAsnProTyrAspMetGln 106
Db 637 CAC-----TTGAAAAGCGTGTACCCCTTATGATGATGATGATGATGATGATGATGATGAT 678
Qy 107 SerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
Db 679 GCTTATATCAATTAGGAATGCTTGAAGAG-----GACGCGATT 720
Qy 127 TyrPheAspTyrProSerAlaGlnAlaGlyTyrGlnGlyArgValAlaIleGln--- 145
Db 721 TTT-----GACAGAGCTGTGCGAGCTTATCTTGTGCTTATGATGATGATGATGATGAT 768
Qy 146 -----AsnMetAlaTyrIleTyrTyrHisGln----- 154
Db 769 AATATGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 828
Qy 155 -----TyrGlnAlaAla-----LysSerProThrLysAspAspTyr 166
Db 829 CTGGCCATGATATCTTACGAGAGAGGCTATGATGATGATGATGATGATGATGATGATGATGAT 888
Qy 167 AsnAsnAlaLysSerAlaLeuGln 174
Db 889 TGCAACTGACCAATATGCTCTCAAA 912

RESULT 17
BC014434 3327 bp mRNA linear PRI 19-SEP-2001
LOCUS Homo sapiens, O-linked N-acetylglucosamine (GlcNAc) transferase
DEFINITION (UDP-N-acetylglucosamine:polypeptide-N-acetylglucosaminyl
transferase), clone MGC:22921 IMAGE:4865031, mRNA, complete cds.
ACCESSION BC014434
VERSION BC014434.1 GI:15680174
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,


```

gene
CDS
/tissue_type="liver"
1..4035
/gene="OGT"
312..3422
/feature="OGT"
/feature="contains TPR motifs: cytosolic uridine
diphospho-N-acetylglucosaminyltransferase"
b-N-acetylglucosaminyltransferase"
/codon_start=1
/product="O-GlcNAc transferase, p110 subunit"
/protein_id="AAC53121.1"
/db_xref="GI:1931579"
/translation="MASSVGNVADSTGIAELAHREYQAGDEFAERHOMOLROEPN
TGVLLSSIHPOCRRLDRSAHFTLAKONPLAASNGYKEREQOLAEHY
RHALRKDFIDGYINLAALVAGDEGAQVAVSALQVNDYCVASDGLNLLKL
GRLEBAKCYKAIETQPNFVAVNSNOCVNAQGEIVLAIHHEKATLDPNLDY
INLGNVLKEARIIPRAVAAYRALISLSPNHAIVHGNLCVYEGQLDLADTYRAI
ELOPHFPDVCNLANALKEGSAVAEDCYNTARLCTHADSINLISPTPADYASNG
EAVRLYRLEVFEPFAASNSIASVLOOQKLEALMHTVEARIRISPTPADYASNG
NTLKMDOVQALOCYTRATQINPAPADNASIHKDSNIPBAIASYTKLRP
DPPAYCNLAHLOVCDWTYDERMKLVSI VAOLEKNEPLSVHPHSLYPLSHG
FRKAIARHGNCLDKINVLKPYEHPKDLSDGRLVQVYSSDFGNHPTSLHMS
IPGMNPKPEVFCYALSPDDGTFNRVYMAEAMHFDLSQIPCGKAADRIDGDIH
ILVNNGYTKARNEFLAPAPLOAMVLGPGTSGALFMFYITDQDTPAEVAYQY
SEKLAMHTFFIGDHANMPHKKKAVIDKRSNGHIDNRIVINGIDLKAFSLSD
VKIYMKCPDGDNDATTNTLNNPVPIMPNTAIVEMIRGQITINFSINCL
ATQIINNCAATGEEVPTITVITSQVGLPEDPAIVGYEYKIDPSITLQMNILK
RVPNSVLMILFPAVGEBNIOQVAMNGLPQNR1IFSPVAREKHVRGQADVCLDT
PLCNGHTTGMVUWAGTPEMTVPGETLARSQSLCTGLCLIAKRSKQYEDIAVK
LGTDEYTKIRGKWKQRISSPLENTKQYTMELRYLYQWHEHYAAGNKDHWIKFV
EVTESA"
BASE COUNT      1113 a      876 c      928 g      1119 t      3 others
ORIGIN
Alignment Scores:
Pred. NO.:      0.0361      Length:      4039
Score:          117.50      Matches:      48
Percent Similarity: 43.09%      Conservative: 33
Best Local Similarity: 25.53%      Mismatches: 68
Query Match:    11.27%      Indels:      39
DB:              10      Gaps:        7
US-10-069-544-2 (1-204) x RNUT6557 (1-4039)
QY      10 ValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProProLysAsn 29
Db      678 TTGGTAGCAGCAGGTGACATGAGGAGCAGCATGCTGCTCTTCAGTAC 737
QY      30 AsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
Db      738 AATCTGATTGTACTGTGCTGTCAGTACCTGGGGAACCTGCTCAAGCCCTGGGTGC 797
QY      50 LeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
Db      798 TTGGAAGAAGCCAGCAGCATGTTATTGAAGCAATTGAGCAGCAACCAATCTTCTGTA 857
QY      70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAsp-----AlaThrHis 86
Db      858 GCGCTGGAGTAATCTGGCTGTGTCTTCAATGACCAAGGAGGATTTGGCTGGCAATTAT 917
QY      87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGln 106
Db      918 CAC-----TTTGAAGAAGCTGTGACACCTTGACCCCAATTTTCTGAT 959
QY      107 SerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
Db      960 GCTATATCAATTTAGGAATGTTTGAAGAG-----GCACGAGTT 1001
QY      127 TyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArgValAlaAlaIleGlu--- 145
Db      1002 TTT-----GACAGAGCTGTGGAGCTTATCTTGGCTTAAAGTTTGAGCCCA 1049
QY      146 -----AsnMetAlaTyrIleTyrTyrGlnGln----- 154

```

```

Db      1050 AATCATCGGTGGTACACGGCAACCTGGCTTGTGTGACTATAGCAAGCCTTATAGAC 1109
QY      155 -----TyrGlnAlaAla-----LysSerProThrLysAspAspTyr 166
Db      1110 CTGGCATTGATATACCTACAGGAGGAGCTATAGAACTGACGCTCATTTCCCTGATGATAC 1169
QY      167 AsnAsnAlaLysSerAlaLeuGlu 174
Db      1170 TGCACCTGACCAATGCTCTCAAA 1193
RESULT 19
HSM804396
LOCUS      HSM804396      5505 bp      mRNA      linear      PRI 10-Jul-2002
DEFINITION      Homo sapiens mRNA; cDNA DKFZp451K0919 (from clone DKFZp451K0919)...
ACCESSION      AL833085
VERSION      AL833085.1 GI:21733676
KEYWORDS
SOURCE
ORGANISM      human.
             Homo sapiens
             Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 5505)
AUTHORS      Bloecher, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B., Amid, C.
and Wiemann, S.
TITLE      Direct Submission
JOURNAL
COMMENT
This clone (DKFZp451K0919) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
source
1..5505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp451K0919"
/tissue_type="CDNA-collection"
/clone_lib="451 (synonym: hlcc1). Vector psport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
5463..5468
polyA_signal
polyA_site
5485
BASE COUNT      1529 a      1124 c      1203 g      1649 t
ORIGIN
Alignment Scores:
Pred. NO.:      0.0537      Length:      5505
Score:          117.50      Matches:      48
Percent Similarity: 43.09%      Conservative: 33
Best Local Similarity: 25.53%      Mismatches: 68
Query Match:    11.27%      Indels:      39
DB:              9      Gaps:        7
US-10-069-544-2 (1-204) x HSM804396 (1-5505)
QY      10 ValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProProLysAsn 29
Db      642 TTGGTAGCAGCAGGTGACATGAGGAGCAGTACACCTTACGCTTCTTCAGTAC 701
QY      30 AsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
Db      702 AATCTGATTGTACTGTGCTGTCAGTACCTGGGGAACCTGCTCAAGCCCTGGGTGC 761
QY      50 LeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
Db      762 TTGGAAGAAGCCAGCAGCATGTTATTGAAGCAATTGAGCAGCAACCAATTTGACGTA 821

```

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	PUBMED	REFERENCE
Neisseria.	1 (bases 1 to 329861)				
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C., Klee,S.R., Morelli,G., Baeham,D., Brown,D., Chillingworth,T., Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N., Holtroyd,S., Jagsels,K., Leather,S., Moule,S., Mungall,K., Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M., Skelton,J., Whitehead,S., Spratt,B.G. and Barrrell,B.G.		Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491	Nature 404 (6777), 502-506 (2000)		
Neisseria	2 (bases 1 to 329861)				
Parkhill,J.	Direct Submission				
Submitted (30-MAR-2000)	Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk				
Notes:	Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).				
COMMENT	FEATURES	SOURCE	gene	CDS	gene
	location/Qualifiers	1..329861	/organism="Neisseria meningitidis Z2491"	/strain="Z2491"	/db_xref="taxon:122587"
		/db_xref="taxon:122587"	/note="serogroup: A"	194..565	/gene="NMA1439"
				194..565	/gene="NMA1439"
				194..565	/note="NMA1439, possible lipoprotein, len: 123 aa; unknown, containing a probable N-terminal signal sequence and an appropriately positioned P800013 Prokaryotic membrane lipoprotein lipid attachment site"
				/codon_start=1	/transl_table=1
				/product="putative lipoprotein"	/protein_id="CAB84676.1"
				/db_xref="GI:7380092"	/db_xref="SPTRMBL:Q9J092"
				/translation="MNKTLSTLPVALILGGCAAGGNTFGSLDGTMGGSIVKAVE	SOCALAEATKRSEMRLLTAMSAEQEMENKICACVAQEAAPNQLTGVDMQLPSTR
				SOCAALATKTAVSCFGLHVR"	212..244
				/gene="NMA1439"	/note="PSO0013 Prokaryotic membrane lipoprotein lipid attachment site"
				complement(571..580)	/label=DUS
				592..601	/note="Core DNA uptake sequence: gccgcctcgaa"
				/label=DUS	/note="Core DNA uptake sequence: gccgcctcgaa"
				/label=DUS	/label=DUS
				638..2047	/gene="chrc"
				638..2047	/gene="chrc"
				/EC_number="4.2.99.2"	/note="NMA1440, chrc, probable threonine synthase, len: 485 aa; similar to many e.g. SW:THRC_MERTL (EMBL:D14071), chrc, Methylobacillus glycogenes threonine synthase (BC 4.2.99.2) (475 aa), fasta scores: E(): 0, 61.9% identity in 475 aa overlap, contains Pfam match to entry PF00291 S-1 dehydratase, Pyridoxal-phosphate dependent enzymes"
				/codon_start=1	/transl_table=1
				/product="putative threonine synthase"	/protein_id="CAB84676.1"
				/db_xref="GI:7380093"	/db_xref="SPTRMBL:Q9J091"

```
misc_feature
875..1780
/gene="chrc"
/note="Pfam match to entry PF00291 S.T.dehydratase, Pyridoxal-phosphate dependent enzymes, score 21.80, E-value 1e-65"
complement(922..931)
/note="Core DNA uptake sequence: gccgctcgttaa"
/label=DUS
2094..2888
/gene="NMA1441"
2094..2888
/gene="NMA1441"
/note="NMA1441, len: 264 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1441"
/protein_id="CAB84677.1"
/db_xref="GI:7380094"
/db_xref="SPTRMBL:OJU90"
/translation="MTVPMQSTHPSLLDLAAKQAAPIA PHILADGTKVIRKAGNH
NARRRYALGNVARYLYKGVLKYKPYPSGSGEPALITSEKRILEYLRPAIGIAPPELLAYRK
NALMFNLNEGIPLDQTIOBEAKGADAMLAGLARVKHKKDFSOQAFRRNKMSDYG
KNISFLBEDDEPSEVLTIACQARDMLCYHTSLILRNKGDSLEAAAEKVGYSIDPP
AEIQIKLAATYKPLPIRIRREHPRWGDARLAASISLSLADMP"
2904..2937
/note="stem loop containing DNA uptake sequences: acgcac
gccgctcgtaa gccc ttccagacgc atata"
2909..2918
/note="Core DNA uptake sequence: gccgctcgtaa"
/label=DUS
complement(2923..2932)
/note="Core DNA uptake sequence: gccgctcgtaa"
/misc_feature
RBS
3065..3068
gene
3076..3852
/gene="fpr"
3076..3852
/gene="fpr"
/EC_number="1.18.1.2"
/note="NMA1442, fpr, probable ferredoxin--NADP reductase,
len: 258 aa; similar to many e.g. SW:FENR_AZOVI
(EMBL:I36319), fpr, Azotobacter vinelandii
ferredoxin--NADP reductase (EC 1.18.1.2) (257 aa), fasta
scores: E(): 0, 69.88 identity in 255 aa overlap. Similar
to NMA1664, fasta scores: E(): 1.3e-30, 34.6% identity in
257 aa overlap. Contains Pfam match to entry PF00175
oxidored fad, Oxidoreductase FAD/NAD-binding domain"
/codon_start=1
/transl_table=11
/product="putative ferredoxin--NADP reductase"
/protein_id="CAB84678.1"
/db_xref="GI:7380095"
/db_xref="SPTRMBL:OJURE3"
/translation="MAAPNQIVLSAHMNTDAYPTPTCRDESILRFENGOFVMWGLNVF
DGKRLAVAYSANWEHLLEFPISIKYODGSLRIQLHVRGVGVLI SKKPTGLTVLG
DLNRGKLHYLSSTGCIAPRLSTIKDEPIEQRKITLVNGVARKYKDIAAYDRPTKEL
PEHHITGLVVEKULIYPIVSREEFHNGRLTDMWGSKLFEDIGLPKINPODRAML
CGSPAMKDKTCVDVDFGLTVSPKTVGRGYLLIERAFVDD"
3394..3768
/gene="fpr"
/note="Pfam match to entry PF00175 oxidored_fad,
Oxidoreductase FAD/NAD-binding domain, score 20.00,
E-value 8.4e-05"
3866..3875
misc_feature
```

```

/note="Core DNA uptake sequence: gccgcctcgaa"
/label=DUS
repeat_unit
3964..3983
/note=">= 90% match to ATTCNNNNNNNGGAAAT"
/label=dRS3
complement(3984..4038)
repeat_unit
/note="RS103"
/label=RS103
4039..4058
/note=">= 90% match to ATTCNNNNNNNGGAAAT"
/label=dRS3
4159..4320
/gene="NMA1443"
4159..4320
/note="NMA1443"
/note="NMA1443, len: 53 aa, unknown, lies within a region
of unusually low GC content"
/codon_start=1
/transl_table=1
/product="hypothetical protein NMA1443"
/protein_id="CAB84679.1"
/db_xref="GI:7380096"
/translacion="MIYKKNCPNKKCYAENKMKWFIENGCSATVIRIGIWKIHH
IDILIPFIA"
complement(4509..6980)
/gene="NMA1444"
complement(4509..6980)
/note="NMA1444, probable P-type cation-transporting
ATPase, len: 823 aa; similar to many e.g. SW:COBA_ENTHR
(EMBL:U13292). COBA, Enterococcus hirae
copper/potassium-transporting ATPase A (EC 3.6.1.36) (727
aa), fasta scores: E(): 0, 34.0% identity in 744 aa
overlap. Similar to NMA1539, fasta scores: E(): 0, 34.4%
identity in 735 aa overlap. Contains hydrophobic, probable
membrane-spanning regions. Contains two Pfam matches to
entry PF00122 E1-E2 ATPase, E1-E2 ATPases and P800154
E1-E2 ATPases phosphorylation site"
/codon_start=1
/transl_table=1
/product="putative P-type cation-transporting ATPase"
/protein_id="CAB84680.1"
/db_xref="GI:7380097"
/db_xref="SPTREMBL:Q9J089"

Alignment Scores:
Pred. No.: 10.2 Length: 329861
Score: 117.50 Matches: 43
Percent Similarity: 46.70% Conservative: 42
Best Local Similarity: 23.63% Mismatches: 78
Query Match: 11.27% Indels: 19
DB: 1 Gaps: 5

US-10-069-544-2 (1-204) x NMA522491 (1-329861)

Qy 4 ArgVallystrpPromeValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThr 23
Db 85796 CGAATCTCTTATTATTAATCTGCTGCTGGCC-----GGCTGGAGCACTTCC 85843
24 -----ProIleProPlyAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAla 41
Db 85844 TACCGCCCTCGCGGGGAGAAAAAGCCATCAAGTTTCAATATCAAAACCAAGTTGGCA 85903
Qy 42 IleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaAlaLeu 61
Db 85904 ATGGAATATATGCGGCTCAGAGACTACCGTCAGGAGCGGCAAGATATGAAGAGCCCTTG 85963
Qy 62 SerIlaAspArgGlnPheAlaProAlaIlyrArgThrLeuAlaLysValIlyrGlnAlaSer 81
Db 85964 AAATCAGACCTTAACAGACCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 86023
Qy 82 GluAspAlaThrHisGlnThrIlysalGlnArgLeuPheGluIlyAlaIleGluLeuAsn 101

```

```

Db 86024 AAA-----GTTACGACAGCGGCGAGAAAGTTTCGGGAGAGCCCTTCATCAAA 86074
Qy 102 ProIlyAspMetGlnSerIlyrMetAspIlyrGlyPheIlyrLeuVal---GlnMetGlyAsp 120
Db 86075 CCCGACAGCTCCGCAACCAACCAACTACGCGCTTCCTGCGGAGAGCTCAACCCG 86134
Qy 121 LeuSerGlyAlaLeuIleIlyrPheAspIlyrProSerArgAlaIleGlyIlyrGlnIlyrArg 140
Db 86135 CTTGCGCAATCTATGCGATATTTGACAAAGCCCTGCGGAGAGCCCACTTACCGACCCCT 86194
Qy 141 ValValAlaIleGluAsnMetAlaIlyrIleIlyrThrHisGlnIlyrGlnAlaAlaLysSer 160
Db 86195 TATATTCGCAACCGAAT-----AAAGCATATGACG 86227
Qy 161 ProThrIlyAspIlyrIlyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGly 180
Db 86228 GCAAACAGGCGGCAATTCGATTGCGGAGAGCCATTTGAAGCTTCCTGCGGAGCCAG 86287
Qy 181 ThrGln 182
Db 86288 CCGCAG 86293

RESULT 22
AX044032 349980 bp DNA linear PAT 24-NOV-2000
LOCUS
DEFINITION Sequence 111 from Patent WO0066791.
ACCESSION AX044032
VERSION AX044032.1 GI:11342916
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Neisseria meningitidis.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 349980)
Piazza,M., Hickey,E., Peterson,J., Tettelin,H., Venter,J.C.,
Messigman,V., Galeotti,C., Mora,M., Ratti,G., Scarselli,M.,
Scariato,V., Rappunoli,R., Frazer,C.M. and Grandi,G.
Neisseria genomic sequences and methods of their use
Patent: WO 0066791-A 111 09-NOV-2000;
CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1..349980
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
/note="Sequence too long, cut in 8 pieces.-seq 1: 1 to
349980 349980 bases-seq 108: 300001 to 649980 349980
bases-seq 109: 600001 to 949980 349980 bases-seq 110:
900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
2272325 172325 bases"

BASE COUNT 86771 a 92803 c 86340 g 84066 t
ORIGIN

Alignment Scores:
Pred. No.: 11.1 Length: 349980
Score: 117.50 Matches: 43
Percent Similarity: 46.70% Conservative: 42
Best Local Similarity: 23.63% Mismatches: 78
Query Match: 11.27% Indels: 19
DB: 6 Gaps: 5

US-10-069-544-2 (1-204) x AX044032 (1-349980)

Qy 4 ArgVallystrpPromeValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThr 23
Db 12661 CGAATCTCTTATTATTAATCTGCTGCTGGCC-----GGTCCAGACTTCC 12678
Qy 24 -----ProIleProPlyAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAla 41
Db 12679 TACCGCCCTCGCGGGGAGAAAAAGCCATCAAGTTTCAATATCAAAACCAAGTTGGCA 126768

```

```

Qy 42 115erleuenuapMetGlyLyLeuaspGlnAlaLysGlnInleuAspAlaAlaLeu 61
    :::::
Db 126769 ATGGAATATATGCGCGGTCAAGACTACCGTCAGCGCAAGCAAGTATTGAAGAGCCCTG 126828
Qy 62 SerAlaapArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSer 81
    :::::
Db 126829 AAATCGAAGCCCTAAACAAAGAGCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 126888
Qy 82 GluAapAlaThrHisGlnThrAlaGlnAlaArgLeuPheGlnValAlaLeuAsn 101
    :::::
Db 126889 AAA-----GTTAACGACAGAGCGCGCAAGAACTTCCGCGCAAGCCCTCTCATCAA 126939
Qy 102 ProLyAapMetGlnSerTyrMetAspTyrGlyPheTyrLeuVal---GlnMetGlyAsp 120
    :::::
Db 126940 CCCGACAGTCCGCAATCAACCAACACACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 126999
Qy 121 LeuSerGlyAlaLeuLeuTyrPheAspLysProSerArgAlaAlaGlyTyrGlnGlyArg 140
    :::::
Db 127000 CTTGCGCAATCTAAGGCAATATTGACAAAGCTTGCGCGCAAGCCCACTCCGACCCCT 127059
Qy 141 ValValAlaAlaLeuGlnMetAlaTyrLeuTyrTyrHisGlnTyrGlnAlaAlaLysSer 160
    :::::
Db 127060 TATATTCACCACTGAAAT-----AAAGCATATGCAAGC 127092
Qy 161 ProThrLyAapAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuLeuSerGly 180
    :::::
Db 127093 GCAAAACAGAGGCGCAATTCGAGTTGCGGGAAGCCATTGTAACCTTCCTCGCCGCCAG 127152
Qy 181 ThrGln 182
    :::::
Db 127153 CCGCAG 127158

RESULT 23
AF145622
LOCUS Drosophila melanogaster clone GH04245 BcDNA.GH04245 (BcDNA.GH04245)
DEFINITION AF145622
ACCESSION AF145622.1 GI:5052533
VERSION AF145622.1
KEYWORDS
SOURCE
ORGANISM Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 3161)
Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazer,R.G.,
Butenoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
Hokins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
Li,P., Moshrefi,M., Pacled,U.M., Park,S., Sequerra,A., Sethi,H.,
Shir,E., Svireks,R.R., Weinburg,T. and Celinker,S.E.
Full length Drosophila melanogaster cDNA sequence
Unpublished
2 (bases 1 to 3161)
Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazer,R.G.,
Butenoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
Hokins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
Li,P., Moshrefi,M., Pacled,U.M., Park,S., Sequerra,A., Sethi,H.,
Shir,E., Svireks,R.R., Weinburg,T. and Celinker,S.E.
Direct Submission
Submitted (23-APR-1999) Berkeley Drosophila Genome Project,
University of California Berkeley, Berkeley, CA 94720, USA
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
FEATURES
SOURCE
1. 3161
/organism="Drosophila melanogaster"

```

```

/d_b_xref="taxon:7227"
/clone="GH04245"
1. 3161
/gene="BcDNA.GH04245"
129. >3161
/gene="BcDNA.GH04245"
/codon_start=1
/product="BcDNA.GH04245"
/protein_id="AAB38597.1"
/d_b_xref="GI:5052534"
/d_b_xref="FLYBASE:FBcd0004245"
/translacion="MVEQIRIMSOQSGSHQPSAHILLDQNPSTGSLVVKQND
IOSLSVGLLELAHREYQADVESAHCQMLPDSASTGTGVLISLHFCRRDK
SAOFSTLAKONVLAAYSNLGNVKEKGOLEADYRRAVRLKPDIDGYNLAA
ALVARDMSAVOAYITLQVNDLVCSVDGKALGRLEAEACVYKALETGCG
FVAMNSICVPRAGETILATHPERKATLDPRDADYINGNVKEKRIPRAYVA
YKRALUSRNNAVHGNLCVYBQGLIDLDITTRALBLQNPFDAYCNLALAKE
KQGVKAEDCYNTALRLCSNHADSLNNLNLKREQYTEATRLYLKALEFDPFAA
HSNLASVLQOOGFLKELAMLYKALRLOTFADAYBNMONTKELDVSGALOCYTRA
ICINPAFADASHNLASIHKDSGNIPEAIOFRTALRKDPDPAYCNLALICDM
TDYDLMKRLVSLVTEQLKRNRLPSVPHSHMLYPLTHCCRKAIAARHANLCLQYV
LHKRPYNFLKULPTKGRLEIKYLSDPGNHPSHIMOSVGLHDEKVEFPCALSPD
DGTTRHKTSRSENFVDLSQIFONKADKIENDGILHLVNMGTGKARNETFLAR
PAIIOYMLGYPETSGASFMDYITDSVTSPLRLAYSEKLSYMPHLYFTGHKQMF
PHLKERIIVCDKQOOSVNDVNVINATDLSPLVENTDVEIKEVNAOKRVEITHKA
ELPNTIOVSMIATGOVOTSLNGVONGATQTNKKAETGEVONIVITRRQYM
LPDDATVVCNPNOLYKIDPOTLESWEIILKNPKSVLWILRLPAVEONIKTVSDPG
ISPRVIFSNVAKKEHVRGQLADICLDTPLCNKHTTMDVLTGTPTVYTLPEETLA
SRVAASQLATLGCPELIARTREYONIAIRLGTKKKKK"

BASE COUNT 1007 a 591 c 648 g 915 t
ORIGIN

Alignment Scores:
Pred. No.: 0.0376 Length: 3161
Score: 116.00 Matches: 50
Percent Similarity: 43.07% Conservative: 37
Best Local Similarity: 24.75% Mismatches: 83
Query Match: 11.12% Indels: 32
DB: 3 Gaps: 7

US-10-069-544-2 (1-204) x AF145622 (1-3161)
Qy 18 SerAlaCyGlnSerThrProLeuProLyAapAsnProGlnInleuAlaGlnLeuArg 37
    :::::
Db 636 TCTGCCGTTCAACCTTATATTAACAGCGTTACATATATATCTGATCTGTTCGCGCC 695
Qy 38 ThrGlnAlaAlaLeuSerleuenuapMetGlyLyLeuaspGlnAlaLysGlnInleu 57
    :::::
Db 696 AGCGATCTGGAAATCTTCTTAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 755
Qy 58 AspAlaAlaLeuSerAlaapArgGlnPheAlaProAlaTyrArgThrLeuAlaLysVal 77
    :::::
Db 756 CTAAGCAATTAACAAATGCTCTGCGTTCGAGTTCAGAGAGTAACTAGCATGCTGTG 815
Qy 78 TyrGlnAlaSerGlnuap-----AlaThrHisGlnThrLysAlaGlnArgLeuPhe 94
    :::::
Db 816 TTTAAGCGTCAAGGAGAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 857
Qy 95 GluLyAapAlaLeuLeuAsnProLyAapMetGlnSerTyrMetAspTyrGlyPheTyr 114
    :::::
Db 858 GAAAAAGCTGTGACTGTTGATCCCAATTTTGGAGATGATCAATTAATTTGGAAACGTT 917
Qy 115 LeuValGlnMetGlyAapLeuSerGlyAlaLeuLeuTyrPheAspLysProSerArgAla 134
    :::::
Db 918 CTTAAGGAGGACAGAAATTTGATTAAGAGCTGTGCGCGCCCTAT-----TTACGTGCA 968
Qy 135 11leGlyTyrGlnu---GlyArgValAlaAlaLeuGlnAsnMetAlaTyrTyrTyrHis 153
    :::::
Db 969 TTTAATCTATCTCCGAATATGCTGATGTTCAAGCAATTTGGCAATGCTTATCTATGAA 1028
Qy 154 Gln-----TyrGlnAlaAla-----LysSerPro 161
    :::::
Db 1029 CAAGTCTTATGATTAGTATGATACATATAGAGAGGCAATAGAACTGCAACCAAT 1088

```

Qy		162 ThrlySASGAspTYrAAsnSnAlalysSerAlaleuGIunArgAlaleuIleSerTyLthr	181
Dd		1089 TTCCCCGAATCGCAGTCCATCTTGCAAAAAGCTCTCAA-----	1127
Qy		182 GlnHISasGluIlleLyseSerTYraSpLysLeuSerAspTYrltysLeuDeuber	201
Dd		1128 GAAAAAGGGCAGATTAAAGAACGTAGAGACTCTCTTAACACTGCACCTAAGACTAATTCCA	1187
Qy		202 AspyTr	203
Dd		1188 AATCAT	1193
RESULT 24			
LOCUS	AF217788	3560 bp	mRNA linear INV 09-FEB-2000
DEFINITION	Drosophila melanogaster O-glycosyltransferase mRNA, complete cds.		
ACCESSION	AF217788		
VERSION	AF217788.1	GI:6942067	
KEYWORDS			
SOURCE			
ORGANISM	Drosophila melanogaster. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 3560) Rubenstein,D.S., McEwen,D.G. and Peifer,M.A. Drosophila melanogaster O-glycosyltransferase Unpublished 2 (bases 1 to 3560) Rubenstein,D.S., McEwen,D.G. and Peifer,M.A. Direct Submission Submitted (20-DEC-1999) Dermatology, University of North Carolina School of Medicine, Suite3100 Thurston-Bowles Bldg., CB 7287, Chapel Hill, NC 27599-7287, USA Location/Qualifiers 1..3560 /organism="Drosophila melanogaster" /db_xref="taxon:7227" 137_3316 /note="OGT" /codon_start=1 /product="O-glycosyltransferase" /protein_id="AAF32311.1" /db_xref="GI:6942068" /translation="MHVSQTIRINMGSQSQSLPSAAHIILDPNPSTGSLVYKQND IOSLSGLLELRHYDAVDPESEAKFCMDLMRODSTNTGVLLLSITHQCKLRK SAQFETAIKONPVLAELAVSNLGNYEKRGLOELADLYRAVALKDFIDGYINLAA ALVAARDMESAVOAITYITALQYNPLCYRSGDLNKLGRLEBAKCYLEAITCGG FAVASNLGCVENAGEIEWLAIHHBEKVATLDPNFLDIYNLGNLKEARIFEDAVAA YLRLANTSPNNNAVHGNIACVYEEGGLDIDLIDPYRRRIELOPNPAPCNLAALKEA KGVAEADDCYNTAIRLCNSNADSNNIANIKRQGYIEATRILYIALRVFPPPAAA HSNLSAVIQOQGKLEALMHTKREARIPLOTADAYSNNGNLTKELODSGLQCYTWA IOINPAFDAHSNLASIHKDSGNIPEAQSYRTALKPKPDPAVCNLAHLQICOMW TDYDIRMKKLAVISYTEQEKNRLPSEVPHSHMLPLTDCDKALIAAHNAICLEKHVV LHKRXENFLKKLPTKGRARIQYLSDFGNHPTSLMOGVPGLHDRSKVEIFCYALSDD DGTRFKHISESENFFVLISOIPCNGKAADI PDNGHIIIVNMNGYTKGANEI FALRL PAPIOVMMGLGYPGTSGASFMDYITDPDYSPLSLAYOSEKLSYMPHYPIGDHKQNF PHUKERTIIVCDKOOSVVDNVTVINATDLSLVENTDVKEIKEVYVNOKPELTHKVA ELPNTTQLVSMIATGOVQTSINGVAVNQGLAKTTQTNMAAAGEVPONIIVTTRQQA LPDDAVVCNFRNOLAKIDPOTLESVWEILKNVPYKVI LAFPVGEONIKIKTVSDSG ISPDVVISNAAKEEHRRGLADI CDUTPLCGKHTSMVLNVMTGTIPVTLPEETLAA SRVASPOLATGCPPELLARTREEYONAIRIGTKKEYIKALARAVVMARVESPLFDOS QYAGLEKTLFLPMMEKYENGELPHIISAV"		
BASE COUNT	1169 a	648 c	716 g 1027 t
ORIGIN			
Alignment Scores:			
Pred. No.:	0.0437	Length:	3560
Score:	116.00	Matches:	50
Percent Similarity:	43.07%	Conservative:	37
Best Local Similarity:	24.75%	Mismatches:	83
Query Match:	11.12%	Indels:	32

DB: 3 Gaps: 7

US-10-069-544-2 (1-204) x AF217788 (1-3560)

QY 18 SerAlaCysGlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArg 37

Db 644 TCTGCCCTTCACAGCTTATATATTACAGCGCTTACAAATATATCTGATCTGTATTGGTCGC 703

QY 38 ThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnLeu 57

Db 704 AGCATCTCGGAAATCTTTCTTAAGCGCTCGCGTCTGTGGAGAAGACCAAGCGCTTTGAC 763

QY 58 AspAlaIleLeuSerAlaAspArgGlnPheAlaProAlaIleArgThrIleAlaLysVal 77

Db 764 CTAAAGCGAAATTGAACATGTCTCGCTTTGCAAGTTGACAGGATAACTTAAGATGTGTG 823

QY 78 TyrGlnAlaSerGluAsp-----AlaThrIleGlnThrLysAlaGlnArgLeuPhe 94

Db 824 TTTAACGCTCAGGAGAAATTTGGTGGCTATACATAC-----TTT 865

QY 95 GluLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyr 114

Db 866 GAAAGAGCTGACCTTGATATCCCAATTTTGGATCATATCAATAATTGGGAAACGTT 925

QY 115 LeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAla 134

Db 926 CTTAAGGAGGACGAAATATTGTATAGAGCTGTGGCCGCTTAC-----TTACGTGCA 976

QY 135 IleGlyTyrGlu---GlyArgValAlaIleGlnAsnMetAlaTyrIleTyrTyrHis 153

Db 977 TTAAATCTATCTCCGAATAATATGCTGTGATGTTCAAGGCAATTTGGCATGCGCTTATGAA 1036

QY 154 Gln-----TyrGlnAlaAla-----LysSerPro 161

Db 1037 CAAGTCTTATATGATTATAGCTATGCATACATATATAGAGGCAATAGAACTGCAACCAAT 1096

QY 162 ThrLysAspSerTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGlyThr 161

Db 1097 TTCCCGCATGCTGATCTGCAACCTTGCAGAAAGCTCTCATA----- 1135

QY 182 GlnHisAspGlnIleLysLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeuSer 201

Db 1136 GAAAGGAGGCGAGTTAAAGAGCTGAGAGCTGCTATATACCTGACTATAGACTATAGATTGTTCA 1195

QY 202 AspTyr 203

Db 1196 AATCAT 1201

RESULT 25

AE007842 10671 bp DNA linear BCT 27-JUL-2001

LOCUS AE007842

DEFINITION Clostridium acetobutylicum ATCC624 section 330 of 356 of the complete genome.

ACCESSION AE007842 AE001437

VERSION AE007842.1 GI:15026533

KEYWORDS

SOURCE

ORGANISM Clostridium acetobutylicum.

Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.

1 (bases 1 to 10671)

Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q., Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitt,J., Wolf,Y.I., Tatunov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P., Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.

Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum

J. Bacteriol. 183 (16), 4823-4838 (2001)

JOURNAL MEDLINE 21359325

PUBMED 11466286

2 (bases 1 to 10671)

Childress,D., Zeng,Q. and Smith,D.R.

REFERENCE

AUTHORS

TITLE

DIRECT SUBMISSION

JOURNAL Submitted (24-JUL-2001) GTC Sequencing Center Production, Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA

FEATURES
source Location/Qualifiers

1..10671
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
complement (124..951)
/gene="CAC3443"
/note="b1cr"
complement (124..951)
/gene="CAC3443"
/codon_start=1
/transl_table=1
/product="Multidrug-efflux transporter transcription regulator, B1cr"
/protein_id="AAK81373.1"
/db_xref="GI:15026534"
/translacion="MENKYLSTGELAKMLGVTKQTVIYYDKVGLISPAKRGKHRY YLEQADELSILTRNLGVPIIDTLKYLVSNAEGCIEMLGQOQERVILEIQKLRIR KRIEGRSKLKQVGVDFEVRFTIEDKATIMEPCYKDEKSTQMSTISCKNSK ELQIDFQNPICSIITKEALSEDKVDYFGIRPKDIGHIVSRFEKPDGIYASTYH KGSYDTVYLSYKRLIENIKEGYEVCCNAYEVDLSTLTSVSSDEVYKLISIQCV"
1043..2431
/gene="CAC3444"
1043..2431
/gene="CAC3444"
/codon_start=1
/transl_table=1
/product="Probable cation efflux pump (multidrug resistance protein)"
/protein_id="AAK81374.1"
/db_xref="GI:15026535"
/translacion="MLNVINKINQKQDVISIFALSWPSILEQLLETAVOYSSAWG QLAGESAAGVLTATATWLTINSBFGVIGVLSYLSIGAKKFRRAKIAVOIIIT LIGIFPMIGITLISPFLPKMLGAPEIRGNASLYFGIICIPMIFRVALIFGVATRS TGMKTPMKKNTLMNLINIVLITLIGKOTIGISGFRITTYGNGIGVGSALAAAT YTVSGIIMPLALYKNIKVSPEKGEKLNIPMRCAVGVFPVTERVATSLGCVBS LVTRLGLTAFAASIAITAEQAFIPEYGEATASTAGNALBKKKLNQISVITI GMAITITVATGTLFLPDPFMSLFIIDPVVIGSGSSVRLVAISEVPFGLIILEGI FNGVGETKTPVLVSIISMGVRLISTICIVFHLGKSWIMCIADNVREFTILFR FLNGAMKCKLNL"
complement (2570..3013)
/gene="CAC3445"
complement (2570..3013)
/gene="CAC3445"
/codon_start=1
/transl_table=1
/product="Predicted acetyltransferase"
/protein_id="AAK81375.1"
/db_xref="GI:15026536"
/translacion="MKVKKILSNKQFLDLLLADBEQDMIDKLYDGENFALYNDL KSTCVNQDSTYELKNIAITYKKYQCKGYKALVKIIPAYYCKCKTIIVTGDSHL TTPYENGPHISHRVKNPFIQNDHPIDPDGQLIMVYLKIDL"
complement (3023..3931)
/gene="CAC3446"
complement (3023..3931)
/gene="CAC3446"
/codon_start=1
/transl_table=1
/product="Predicted epimerase, PhzC/PhzF homolog"
/protein_id="AAK81376.1"
/db_xref="GI:15026537"
/translacion="MIRKYNLYQIDSFTEKFTGNPAGVITNADGLTDYEMOKIAREL NNSETAFIPSSKDYDAWVFFPTSEVPICGATIAAHARAIEENNLSTRYNHT GAGILPVDVAKENDDYKIMTQKIEFGSVIDGKELRLALNIEKSDLLBNYKIOI VSTGSHVMGIKSLKTLNKQPNYDLSLSTSKIKNGYVPAVTSDESDIILHGM PAPAIGINEDPVTGNANPUGAIVVHKLACLYDINSIRFPAKQGEAIGRGITIEVR VDEKEPVEKISGNAVVVFSEVLN"
complement (4036..4722)
/gene="CAC3447"

CDS complement (4036..4722)
/gene="CAC3447"
/codon_start=1
/transl_table=1
/product="Protein-disulfide isomerases DsbC/DsbG"
/protein_id="AAK81377.1"
/db_xref="GI:15026538"
/translacion="MKIEVSDSFVCPFCYMGKRRLIALNEFEYKTDVEISFKSELD PSKAKYSENIHELIAKKYKISIEQAKASNDQIVLOAQAIGLYNFKLIIPNTPAH RLISOYKTKGKNELSEKIFKAYFVDSILNIDYKLDLAEVGISRDSISRLTESNO YNBEVREDEKRNASKIGIDVPIYFVPDCKAVSQAQPAVEALYKIDEGKVNDS S GSCSGKCDI"
4855..5427
/gene="CAC3448"
4855..5427
/gene="CAC3448"
/codon_start=1
/transl_table=1
/product="Probable teaA-like protease"
/protein_id="AAK81378.1"
/db_xref="GI:15026539"
/translacion="MKLVICIGSLTYGVSSSTSWVELLKAFSINVINKINGDPT SGLHSYSDIVAKKPNYITINAGTNDLMDSTKIKDNIELLEAKAENNTIPLA LQPPVISPAAKYWSEIDYERVALSLDIYIRAKAYCINNIAFSENVIFINKONI BALVNGIHPNSGKLMFMBLSKTIINKTI"
5444..6361
/gene="CAC3449"
5444..6361
/gene="CAC3449"
/codon_start=1
/transl_table=1
/product="TPR-repeat-containing protein"
/protein_id="AAK81379.1"
/db_xref="GI:15026540"
/translacion="MKDNIQIDKYLKKAEBAPIDNKLVSLSLDYKNTVETLNGEDVD T IYNLAIYDSLGEVBAEERYKALSIDYBEAAVGLATTIDKEDYBEALYLNYKA IYNLPIYHAAYFLAAYDLSGQKNIAIETYEKLSLDMDPMSNINLGCITYEONKN DLAVERFSALKNIPNNHALLFMGKVICFPNNI EKAINEYKSGISEKNSYESYANL AVIYKRDDEKGIQILSGIENKTHPIFYNRSCTVALINENKSCBDIITLAKYF EPLKYLDEBELESVRELSFKSFAKAN"
6490..7413
/gene="CAC3450"
6490..7413
/gene="CAC3450"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AAK81380.1"
/db_xref="GI:15026541"
/translacion="MISNOYVRLSLEMLFPLRIYKENVVIGASLPKVPATLMEI LANKTLDMLISKTVALSKGNTSRMANSTLITPLTLSEKVTSLATGVPIATLTS KEISLSGRDYIRGIMVTVAVSLINKALPLPVFVYNFQAGLSKIMSCAASPTTPS NIDHIVREARAVYNNLMLEKQEMDTPQGIIRORI PMNDIMEDHAEFRGLVDSQ TSLFNTNANFVRFFDIEVATSELTNPNLNNTNITNIVSLYTEFNFKSTATVGLA CCKKAMAPLADHVIREANVYRLRSFTS"
7629..9005
/gene="CAC3451"
7629..9005
/gene="CAC3451"
/codon_start=1
/transl_table=1
/product="Sugar/(Na⁺(H⁺)) symporter"
/protein_id="AAK81381.1"
/db_xref="GI:15026542"
/translacion="MIGSFRIKRRKIKIYASGDLASNLIVOTISITLLEFYTVFGIS AGAGAGVLPVVRPIDAINDPIIGTLVDKNTNRGRPRPYLVGAAPFAVAFICFTTP NFATKSLIYAAVTYGLSITCYINVPYGLATSDNOEIVSTSVSPFANLGG VTSYVPLVLSAFTSGSGGMOITWSLIGAGAFLFFCSTSRKRYKSVNOCHK IKSQDLFQPKNRPILIVSIFPVLLFGINSINSIGIYITVNVRADIVQVTVIG SIPAPVCPLIIPKINKIGKPLISSLITLTVGTSLSLVITPHAAVALIVSVITSI GSUTAGAFMSLIPETIEYEGYTGRLSGLYAIIGFFKCGNALGAAPVIGIILGF GYVANTQTPHALTGILTAIVVPAVLMILALIDITFYNLDDKKYHIIATLKERAL NRGENLNL"

Db	6862	GCTAACGACGCAAGTACCA-----	ACTGAGATGGCAGAGCG	6903	source	1..13991 /organism="Methanosarcina acetivorans C2A" /strain="C2A" /db_xref="taxon:188937"
Qy	37	ArgThrGlnIleAlaIleSerLeuAspMetGlyLysLeuAspGlnAlaLysGln	56	gene	200..538 /gene="MA1606"	
Db	6904	CGCATGCGCTTGTTAGGCTATTGAAATGCGACATGATCAAGACGGGAAAT	6963	gene	200..538 /gene="MA1606"	
Qy	57	LeuAspAlaIleLeuSerAlaAspArgGlnPheAlaProAlaTyr-----	Arg 72	CDS	200..538 /gene="MA1606" /codon_start=1 /transl_table=11 /product="predicted protein"	
Db	6964	TTAGAAAAAGCACTA-----	CAGCATGGCGCGGATTTATTCGGCTCACACTT	7011	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"
Qy	73	ThrIleuAlaLysValTyrGlnAlaSerGlnAspAlaThrIleGlnThrLysAlaGln	92	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Db	7012	TCATGCGCGCACTATTATGAGCT-----	GTTGGTGAGAAATGATTCGCGGGAA	7062	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"
Qy	93	LeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly	112	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Db	7063	ATGTACCCGACGCGCACTAAGCGACGCAAAACGGTAAATGCTGAATACCTCGCT	7122	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Qy	113	PheTyrIleuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSer	132	CDS	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Db	7123	ACCTTTGTGTAGCAAGCGAGATGTATGACACGCGCATCAATCTTGGCGCT-----	7176	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Qy	133	ArgAlaIleGlyTyrGlnGlyArgValAlaIleGlnAsn--	MetAlaTyrIleTyr 151	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Db	7177	-----	GGGTCGACCACTTATTACTTAAAT	7206	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"
Qy	152	TyrHisGlnTyrGln-----	AlaAlaLysSerProThrLysAspAspTyr 166	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Db	7207	TCAGCAGCTTACGACAGCGCGATGTGTCTCAAGACGGGAAACGAC-----	7260	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Qy	167	AsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGlyThrGlnHisAspGluIle	186	CDS	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Db	7261	---AACGCTCGTGAATATTTCAACGCTATC-----	GATCAGCATCA--- 7302	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Qy	187	LysLysSerTyrAspLysLeuLeuSerAspTyrLysLeu 199		gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
Db	7303	-----	AATCGCTCTTCTGTCATTTTACAGCTG 7329	gene	557..889 /transl_table=11 /transl_table=11 /product="predicted protein"	
RESULT 27	AE010832/C	AE010832	13991 bp	DNA	linear	BCT 03-APR-2002
LOCUS	AE010832	13991 bp	DNA	linear	BCT 03-APR-2002	
DEFINITION	Methanosarcina acetivorans str. C2A, section 177 of 534 of the complete genome.					
ACCESSION	AE010832	AE010299				
VERSION	AE010832.1	GI:19915481				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
JOURNAL						
FEATURES						

/protein_id="BAB60156.1"
 /db_xref="GI:14325252"
 /translation="MVDMAKDEFDMDNDEFKDFKDFGIDIKSLNNRMRIRRL
 NDPIITSEPPVGYGTIRIGSDGKPIFOEFNVGIRITGVAPIRGRVPEPTDINDDD
 KKVLTPELPVSKESIDIKISEVNTLNVEDQKRYKSIDPDKLPETVAKFNN
 GLDPTVEKAKSKESGGRKSIQ"
 complement(8303..11350)
 /gene="TVG1035393"
 complement(8303..11350)
 /gene="TVG1035393"
 /gene="TVG1035393"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="membrane protein"
 /protein_id="BAB60157.1"
 /db_xref="GI:14325253"
 /translation="MFESTFKGIGLFSARHSTAVIISWILLVLAPAPSLPHTNY
 NIASDIPTNPANIASNLQTSFNTSDSFVITNDTSDIDQSLTALNMCEAIM
 SYLSEGSVAVSSILVYENNTLRVSLSTGELKSTIDLNKLTENSVTLTKTLDIT
 ANPLFLPATIYSTLTITRYKSGDLNSEHAIYNTVDNRTLENYEELIYNYNG
 FSLNNTLNSSNIHEVDLVNANETVGIAGKAGFPFLISVDHNLTRTFNVK

Alignment Scores:

Pred. No.: 21.3 Length: 279160
 Score: 113.50 Matches: 44
 Percent Similarity: 40.45% Conservative: 28
 Best Local Similarity: 24.72% Mismatches: 67
 Query Match: 10.88% Indels: 39
 Gaps: 6

US-10-069-544-2 (1-204) x AP000995 (1-279160)

QY 28 Lysanahnpoglnleualglnleargthrglnleialleserleuaspmet 47
 |||||
 DB 236763 AAGAACATCCCGATTATTTCCCTTACG-----ACCATATGTTATGTTCAATG 236813
 |||||
 QY 48 Glylsleuapglinalyeglnglnleuaspalaialeuseralaasparglnphe 67
 |||||
 DB 236814 GGTGATTTTAAGAGGTCAAAAGAGAGCCGACCAAGCATMAAGGACGATCAAGTTAT 236873
 |||||
 QY 68 AlaProalaTyArgThrLeuAlaLysValTyrglnalasergluaspaIathrhlsln 87
 |||||
 DB 236874 CCCTTGGCTTACTTAACCGGACGCTGCACACAGGGGCTTGAGATGTT----- 236924
 |||||
 QY 88 ThrlysaIaglnaagleupheglulysalallegluleuansprolysaspmetclnsr 107
 |||||
 DB 236925 GATGAGAGGAGAAAGATTTTCGAAATAATACATGAGAGCCAGTCCAGATCCT 236984
 |||||
 QY 108 TyrmelaspTyrglyPheTyrlenuValGlnmetGlyaspLeuserGlyAlaLeuileTy 127
 |||||
 DB 236985 CACTACGATCTC-----GCGAATATAC 237008
 |||||
 QY 128 PheaspTyrsProserArgalaileGlyTyrglnuglyArgValAlaAlailegluasmet 147
 |||||
 DB 237009 TTTGAAAAAAAAGAT----- 237023
 |||||
 QY 148 AlaTyrlleTyTyThsGlnTyrglnAlaAlaLysSerProThrylsaspAspTyrsn 167
 |||||
 DB 237024 -----TATTAAGAGCTATGAGAGGTAAAGAGTCTTAAGAAATGATAGAGC 237074
 |||||
 QY 168 AsnalalyseeralaleuGlualaleuileserGlyThrglnhsaspGluileLys 187
 |||||
 DB 237075 AATTAAGAGCTATGAGAGCTCAATTCATCTTGAGCAAAATGAC-----GGA 237128
 |||||
 QY 188 LysSerTyraSpLySLeuLeuserAspTyrlYsLeuLeuserAspTyrlYs 204
 |||||
 DB 237129 GAGGGTAACTCAAGACCTTCTTGAGAGCTTCAAAAACAGAGAAATTCAAA 237182
 |||||

RESULT 30
 SCYBL084C/C 2900 bp DNA linear PLN 11-AUG-1997
 LOCUS SCYBL084C
 DEFINITION S.cerevisiae chromosome II reading frame ORF YBL084c.
 ACCESSION Z35845 Y13134

VERSION 235845.1 GI:536135
 KEYWORDS
 SOURCE
 ORGANISM
 Saccharomyces cerevisiae.
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
 AUTHORS
 Domdey,H., Gassenhuber,H., Obermaier,B. and Piravandi,E.
 JOURNAL
 Unpublished
 2 (bases 1 to 2900)
 TITLE
 MIPS.
 JOURNAL
 Submitted (30-AUG-1994) Data collected by MIPS on behalf of the
 European yeast chromosome II sequencing project. MIPS at the
 Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
 Martinsried, FRG; E-mail: Mewes@mips.embl.net.org

REFERENCE
 AUTHORS
 Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C.,
 Barthe,C., Baur,A., Becam,A.M., Bileau,N., Boles,E., Brandt,T.,
 Brendel,M., Brueckner,M., Busserreau,F., Christensen,C.,
 Contreras,R., Crouzet,M., Czilepuch,C., Demolis,N., Delaveau,T.,
 Daignon,F., Domdey,H., Ducrestius,S., Dubois,E., Dujon,B., El
 Bakoury,M., Enlian,K.D., Feuermann,M., Fiers,W., Fobo,G.M.,
 Fritze,C., Gassenhuber,H., Giansdorff,N., Goffeau,A., Grivell,L.A.,
 de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K.,
 Jacy,C., Jacquet,M., Jauniaux,D.C., Jauniaux,J.L., Kallio,T.,
 Kleau,A.J.E., Louis,E.J., Li,Z.Y., Maat,M.J., Maillet,L.,
 Mannhaupt,G., Messenguy,F., Miosga,T., Moilemans,F., Mueller,S.,
 Nasr,F., Obermaier,B., Perae,J., Pierard,A., Piravandi,E.,
 Pohl,F.M., Pohl,T.M., Potier,S., Profi,M., Purnelle,B., Ramezani
 Rad,M., Rieger,M., Rose,M., Schaff-Gersenschlager,I.,
 Scherens,B., Schwarzlose,C., Skala,J., Stomski,P.P.,
 Smits,P.H.M., Souciet,J.L., Steensma,H.Y., Stucka,R.,
 Urrutiarazu,A., van der Aart,Q.J., van Dyck,L., Vassaretti,A.,
 Vetter,I., Vierendeels,F., Visser,S., Wagner,G., de Wergifosse,P.,
 Wolfe,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kleine,K.
 Complete DNA sequence of yeast chromosome II
 EMBL J. 13 (24), 5795-5809 (1994)
 95112788
 7813418

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 FEATURES
 source
 Location/Qualifiers
 1..2900
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /chromosome="II"
 complement(354..2630)
 /gene="CDC27"
 complement(354..2630)
 /gene="CDC27"
 /note="ORF YBL084c"
 /codon_start=1
 /protein_id="CAA84905.1"
 /db_xref="GI:536136"
 /translation="MAVNPBLAPRLTSLRSPEDDAATSTIIOODCOALIOOLNYS
 TABPRLAEILYACSLIDKSSVYWSDAVYIYALSLFINKSYTHAFQSKFKYHUGIA
 YIFGRALQDSQGVNEALITLISLIVNESNSNTRINNVNLNVHLPDALTNLL
 GNLVYMLDLSKSGAFHSEALAINPILYKSMRAVNLKRVFDPIDAKKSNH
 NNNAASSPSTSLHSEPRSPSLYKTKNNNNINNNVTLFOSNSPPSTASAF
 SSIQHSRSQOQOANNTSIRTCOMKNTQPKNAKNSKSSALPNINISMLVSPSSKOP
 TTSIAKYNNRKLITTPSKLTLNDRHNNNNNNNNNNNNNNNNNNNNNNNNNNNN
 TTFTKPRNYSSTGRITTSKPKRSLISNSLITSDYQITLPEIRNPLIIRSSQY
 NSPKALRLEPSQIPSHIKOTWPCLVQDKHFEITINDMSLKYNRLKIDLPARVQ
 MEIFSTLWHHDKYSSNLANGMDTWTWCCGCGNLISLQKHDAALKAPEKA
 TQIDPNFAVAYTLOGHESNSDSAKCYKALACDPQHNAYAYAGTSSAMKGOY
 EERTLYFEKARSINPNVYVILCCGSGLEKLGKRYKALQYVLAQHLQPTSLSTYKM
 GGLYSMTIRYNVALQTFEELVGLVDPDAHAHLYLQCTYRIVGRKKDAIHELTVANMD
 PKNQVYIDELQKCHQOE"

BASE COUNT 900 a 485 c 542 g 973 t
 ORIGIN
 Alignment Scores:

Pred. No.: 0.0683 Length: 2900
 Score: 113.00 Matches: 46
 Percent Similarity: 46.11% Conservative: 43
 Best Local Similarity: 23.83% Mismatches: 66
 Query Match: 10.83% Indels: 38
 DB: 8 Gaps: 9

US-10-069-544-2 (1-204) x SCYBL084C (1-2900)

QY 38 ThGlnleAlaIleSerLeuLeuAsp----- 46
 DB 1055 TCAATTGGCAATGGCTTAATGATACAAATGCTAATAGCCGGAACATGCTGTGT 996
 QY 47 MetGlyLeuSerLeuSerGlnAlaLeuGlnLeuSerAla----- 63
 DB 995 ATAGCTAATTTGCTATGATGCAAAAGATCATGATCCCAATTAAGCTTCGAAAA 936
 QY 64 -----AspArgGlnPheAlaPheAlaTyrArgThrLeuAlaValTyrGln 79
 DB 935 GCTACTGAGTTAGACCCAAATTTTGATACGCGAT---ACTTGCAAGGTCATGAACAT 879
 QY 80 AlaSerGlnAspAlaThrHisGlnThrLeuAlaGlnArgLeuPheGlyValAlaIleGln 99
 DB 878 TCTTCCACGATCTTCTGATTCGCCAGACA-----TGTATAGAAAGCGCTAGCT 825
 QY 100 LeuAspProLeuAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGly 119
 DB 824 TGTGATCTCAGCATTAACATGATTAAGGATGATGATGATGATGATGATGATGATGAT 765
 QY 120 AspleuSerGlyAlaLeuIleTyrPheAspLeuProSerArgAlaIleGlyTyrGlnGly 139
 DB 764 CAATATGACAGACGCTGTATATTTGAAAG---CAAGTCATTAATTCGCGTCAAT 708
 QY 140 ArgValVal-----AlaIleGlnAspMetAlaTyr----- 149
 DB 707 GTTGTGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 648
 QY 150 IleTyrTrpHisGlnTyrGlnAlaAlaLeuSerProThrLeuAspAspTyrAsnAspAla 169
 DB 647 CTACATATATATGATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 557
 QY 170 LysSerAlaLeuGlnArgAlaLeuIleSerGlyThrGlnHisAspGlnIleLysSer 189
 DB 596 AATATATAGATGCGCCAGTGTCTCTATTCATGACATATATATGTTGCTTGCAACT 537
 QY 190 TyrAspLysLeuLeuSerAspTyrLysLeuLeuSerAsp 202
 DB 536 TTGGAAGATTGGTG-----AACTGCTTCTGAT 507

RESULT 31
 SCNDCH2 78600 bp DNA linear PLN 12-APR-1996
 LOCUS S.cerevisiae genomic DNA, chromosome II from Y element to IL1
 DEFINITION
 ACCESSION X79489.1 GI:496661
 VERSION X79489
 KEYWORDS IL1 gene; transfer RNA-Leu; transfer RNA-Leu-uaa; transfer RNA-Phe.
 SOURCE Saccharomyces cerevisiae.
 ORGANISM Saccharomyces cerevisiae
 Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 78600)
 Obermaier, B., Gassenhuber, J., Piravandi, E. and Domdey, H.
 Sequence analysis of a 78.6 kb segment of the left end of
 Saccharomycetes cerevisiae chromosome II
 Yeast 11 (11), 1103-1112 (1995)
 JOURNAL MEDLINE
 PUBMED 96076635
 7502586
 2 (bases 1 to 78600)
 Obermaier, B.
 Direct Submission
 JOURNAL Submitted (27-MAY-1994) B. Obermaier, Laboratorium f. Molekulare

FEATURES
 source Biologie, Genzentrum, Am Klopferspitz 18a, 82152 Martinsried, FRG
 Location/Qualifiers
 1..78600
 /organism="Saccharomyces cerevisiae"
 /strain="S 288c"
 /db_xref="taxon:4932"
 /chromosome="2"
 /clone="1", 1001.1/1002.8/1002.10/1003.2"
 /clone_1b="cosmid library H. Feldman, Munich"
 1..115
 /note="Start Y Element"
 379..424
 /bound_molety="OBFL"
 605..615
 /note="ARS consensus"
 1768..1776
 /bound_molety="HAP2/3/4"
 2223..2528
 /gene="YBL0801"
 2223..2528
 /gene="YBL0801"
 /codon_start=1
 /product="C-101 protein"
 /protein_id="CAA55987.1"
 /db_xref="GI:496662"
 /db_xref="SWISS-PROT:P38161"
 /translation="MEMILFLNESYTFPHFRFMSYTLVHMSCVFVCAECGNAYVRGAG
 CLEKPPCAVKKPPSVKXKIRILDDPREANCLSHLVCPKRPCKATSLILIPBG"
 2894..3216
 /note="solo-delta-rem"
 3139..3466
 /note="solo-delta"
 3467..3568
 /note="solo-delta-rem"
 3629..3712
 /gene="YBL0804"
 3629..3712
 /gene="YBL0804"
 /product="tRNA-Leu"
 /codon_start=1
 /product="F-196 protein"
 /protein_id="CAA55988.1"
 /db_xref="GI:496663"
 /db_xref="SWISS-PROT:P38162"
 /translation="WVDRNRTFTAPDSLLETLNLTPEPNDEPSLTITITVTRRCVDPSELI
 DFLRLRHGSDDIIRKLNRYKGSINGKCKEFLKQELVPMQIRNNITISFCKE
 AAEKNEPDOCCKNKKTTAEPIDARIDPYARAPRAEQEAQYDMWTYVTEVANNR
 KLEQILTSTTBGILRONCEONDDYLFKEPFCNDNS"
 complement (4893..7925)
 /gene="YBL0806"
 complement (4893..7925)
 /gene="YBL0806"
 /codon_start=1
 /product="E-1010 protein"
 /protein_id="CAA55989.1"
 /db_xref="GI:498989"
 /db_xref="SWISS-PROT:P38163"
 /translation="MFKSRHLKNVSNATKSARVHDVNSGSKPEFTDKICTYING
 RITVTPDYTOSLAVATTAGRIHVYCGQIEVPTLRKROIKNHRAPIKGIYIADV
 EKSNIIVSVHKKOITTYFCNSITCEFTDPPDLMLIGLSGSLITIDVDRNMSK
 LKTIENQKSVLPKRLSPVISIQMNPIDIGTILISYHITVYSFIDYKVOHPFYQ
 LPIYAPGQSLSTNIEKRPVQISLYHNSLHILVHBDNALVFNVDVNSGKLOHFPYQ
 IFETHPNPALCKDCSFETPAIFVSWLCORNEPYSLLATKATENDPCLOEARTS
 ILUGTPNPNVSPKMSKYAKPVQOKLFSILGKAPLNPPLPKASPYFGGCHDTN
 ILILLEDELETLIYPAGSPSSKASIPRSILAMWRPTTCTIAGOSVOKNLMGMWTI
 AOSBSFLKGCIPATNIRRHETRSALITSHSGSVRIWPAHSBVTNDALVFVNTAKV
 LNRATNLAKNISFASSETLAVSSSEGVVILFKETNKFYQLPESVDNLQLPESFS
 LDRSKTILVDVDRGPTNKGIFSTVTHAKGAVASIANSNIGVAVAFEGTLII
 ILDRGPALIFNENIRIVISKAGSSYSTVAFECVMBYGDGDFSSILMLCGTIDGLMTFK


```
/product="hypothetical membrane protein"
/protein_id="CAC12064.1"
/db_xref="GI:10640250"
/translation="MTCGICGENALLPIPISSKLTLLCVASVMPFRREYAYIRGV
LSAVALLTIADIVSFAVFPSPFALDITFYINAFTGLIFALLITYGLEGSRSL
LSAALLAAYFVLDYRSRLTLRLADYGAALFPMILITLISRTALAFYNNILTRPL
TVVFPFGAASALLAYNGLIHHMPRAHEBQLILYLVAVFSSVAFVHSRQVIO
YESHAS"
complement(5692..6330)
/gene="Ta0936"
complement(5692..6330)
/gene="Ta0936"
/notice="weak similarity to known protein. Relation: similar
to phosphoserine phosphatase PIR:S53931."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC12065.1"
/db_xref="GI:10640251"
/translation="MRKERMIKLAIPDMGTLDPSSMEYVHRLHTDNNMFLY
NGRISYEPRSRDLVAMLRHPRKDKDIDIGLIDIKLPSSSEVNGKKEGLITA
VSGGISLDCILNEMOYDNLNVATLDGGFVQPMGTRVPERKNTAVKSLQSM
LHISKDETISIGDSMENGMIHNSKAVIIGKGRHGEIPLGNIRMLSYI"
complement(6427..7815)
/gene="Ta0937"
complement(6427..7815)
/gene="Ta0937"
/notice="similarity to known protein: phosphoserine
phosphatase (EC 3.1.3.3) - Yeast (Saccharomyces
cerevisiae); PIR:S53931"
/codon_start=1
/transl_table=11
/product="polyI--tRNA synthetase related protein"
/protein_id="CAC12066.1"
/db_xref="GI:10640252"
/translation="MENKENSEMYNEITLAEISDKRYPVKANNVWLPYKMKINSI
IDSIVRAVDKRNQOEVPFPIIGRSMLEVEFEHIRGENEIVWYTGKGEKEEELA
LRPESAMPMYFIMLRSHADLPKTIQIVSVRYRRTKTRSPFIRLREHIFEAHRA
HATYEDAEAMQDQKEIMRETSILCLPYFDORPDMKPKGAWTTLAFDVLPSGSS
LOIGTHOYGTNFSKNDIKYKEDGTFDVAHQTTYGSKELLAAIIGIHDDGGLYI
PVDVAPIQVILIPFGAVMEYARDVENTLNSINIRKVDRENYTGYKENDMEKG
VPLRIEIGEREVKNRITLISMRNIQKLTIERSKLIYEPDTLIRIEKMEENQKVF
KDHVFRATKLEEFNRDGLITAFMGSKESDKIBATEKSVLGHMVDSSETGCVVCG
KNGMAVFSRSY"
complement(8008..8340)
/gene="Ta0938"
complement(8008..8340)
/gene="Ta0938"
/notice="no similarity"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC12067.1"
/db_xref="GI:10640253"
/translation="MSEVNIYVNGREAGSKSGCALCATWGDYHADFLGDLFFCCD
ICAEFPMNMDEAFKHTARHNVDELHIDNQVQLRNVLKNGEGRLEFYVKGFGAVI
KERITRK"
complement(8333..9223)
/gene="Ta0939"
complement(8333..9223)
/gene="Ta0939"
/notice="weak similarity to unknown protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC12068.1"
```

Alignment Scores:

Pred. No.:	29.3	Length:	297750
Score:	112.50	Matches:	40
Percent Similarity:	39.04%	Conservative:	33
Best Local Similarity:	21.39%	Mismatches:	57
Query Match:	10.79%	Indels:	57

```
DB: 1 Gaps: 5
US-10-069-544-2 (1-204) x TACID4 (1-297750)
QY 28 LysasnandProglInleualaglnlleargThrglnllealieserLeuleAspMet 47
Db 196251 AAGAAATATCTGAGCTGCTTGGCTCAAA-----AGCATATGCTACGGGCTCTCG 196301
QY 48 GlyLeuleasporlnalalyserGlnInleualaspAlalaieuserAlaaspGlnPhe 67
Db 196302 GGAAAGTTCAACAGATCAAGATCGAGCGCACCAAGCAATAAAGCTGATCCAACTAT 196361
QY 68 AlaProbalatyraThrleualalyseValtyrglnalaserGlnalalnHieglN 87
Db 196362 CCGTGGCTTATTAACAACAGCGCCACGAACAGACTTGAAG-----TAC 196412
QY 88 ThrlyealaglnargleuphegluLyalaileGlnleuanProlyasapMetGlnSer 107
Db 196413 GATATCCCAAGAAAGATCTGCAGAAATATCTTGAATGCAACCAAGATCCAGAGCC 196472
QY 108 TymerAspTyrglyPheTyryleualGlnMetGlyaspleuserGlyAlaleulleTy 127
Db 196473 TACATGGAT----- 196481
QY 128 PheaspysProserarAlaileGlyTyrgluGlyArgValAlaAlaileGluAspMet 147
Db 196482 -----CTT 196484
QY 148 AlaTyrlleTyryThrglnInTyrglnAlaAlalyserProThrylasapTyraSn 167
Db 196485 GCCGATATGAAATATCAAGAGTGACTACAGAAAGCA-----CTTCAG 196529
QY 168 AsnAlalyserAlaenluarAlaAlaLeulleSerGlyThrglnHisasp----- 184
Db 196530 CACGTAAATCTGCAATAAAGAAAGATPACAGAGCACAGCACGACTTAATTG 196589
QY 185 -----GlnleTyLysSerTyryaspleuserAspTyx 197
Db 196590 AACATCTTCTTGGCGAGAGGATATAGAACTATCTTAAGAGAGCTGAGAGCTTTC 196649
QY 198 LysleuleuserAspTyryLys 204
Db 196650 AAGGATACAGAGGATTTCAAA 196670
RESULT 33
PHY17720 3559 bp mRNA linear PLN 22-JUL-1998
LOCUS PHY17720 Petunia hybrida mRNA for SPY protein.
DEFINITION Petunia hybrida mRNA for SPY protein.
ACCESSION Y17720
VERSION Y17720.1 GI:3319681
KEYWORDS SPINDLY protein; spy gene.
SOURCE Petunia x hybrida.
ORGANISM Petunia x hybrida.
REFERENCE 1 (bases 1 to 3559)
AUTHORS Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
JOURNAL Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE Asteridae; easterids I; Solanales; Solanaceae; Petunia.
JOURNAL Submitted (06-JUL-1998) D. Weiss, The Hebrew University, Faculty of
Agriculture, Dept of Horticulture, P.O.Box 12, Rehovot 76100,
ISRAEL
COMMENT Related entry: U62135.
FEATURES
1. 3559
location/Qualifiers
/organism="Petunia x hybrida"
/cultivar="VA"
/db_xref="taxon:4102"
/tissue_type="anthers"
/clone_lib="lambda ZAP"
```



```

gene      /dev_stage="flowering"
198..2996 /gene="spy"
198..2996 /gene="spy"
198..2996 /codon_start=1
/product="SPINDLY protein"
/protein_id="CAA6834.1"
/db_xref="GI:3319683"
/db_xref="SPTREMBL:O62039"
/translation="WANTKOVENGKESDGLNNGNFKGVQSSDSKGSFVRISPVKK
SFEKDAITYPANILRSNKFVDALALYESVLQDSGIESLIGKGI CLQONMRLAF
ESFEBAIKLPONACALTHGILYKDEGRVBAESYQKALPSYKPAECIALVL
TDIGTSIKLAGNSEOGLQKYELAIKIDSHAPYVYNGVSEMOQDMALNCEKKA
IERPMYBAACNMGVYIKONGDELSAICERCLAASPNBEIKNMMAIITDGLGVK
KLEGDINGVAYIKKALYYMHYADAMNINGVAYGELKRDMAIVPELAFHNPICA
EACNIGLVITKDNLDKAVECTOMALTIKPNFSOSLNIGVYTVQGRDAASMTIE
KAIIPNPTVAKDNNLGLVLRDAGNISLAEVQCKIDPDSRNAGONRLAMNYIN
EGSDDKLYEARNDMGMRFRMLYQYNSMDSKDEPRLVIGYVSPEFTHSVYFIEA
PLAYHDYANKVYISAVVACADAKTNFRDKVKGSGVMDIYIGDEKYSMTREDK
VDIMIELTGHANNKLGMAACRPAPVQVTVIGVNTGLPTIDYRITDSWADPSTKO
KHYBELVRLPDSFLCYGSPSPAGPVPSPAPLTNGPFGSPNNLAKITTPLOVMARI
LCAPHSRLIVKCKPQECDSVRQRFSLIQLEPQVRLVPLILNHDMAYSILM
DISLDTFPYAGTTTTCESLYMGVPCVTMGSAHVANNVSLATVGLRKLVANDEY
VELAIQLASDVTSLNSLRMSLRIMAKSPICDAQOFQONLESTYRSWMRYCDGDVPS
LRMRELQOQOQTLAEVLVPEESPVSPISIEKTRISASKQGPINKNGFVSPALVNSST
IEBNVQNLQAGNPKROS"
BASE COUNT 1052 a 668 c 830 g 1008 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 0.112 Length: 3559
Score: 112.00 Matches: 47
Percent Similarity: 37.80% Conservative: 32
Best Local Similarity: 22.49% Mismatches: 94
Query Match: 10.74% Indels: 36
Gaps: 6
US-10-069-544-2 (1-204) x PHV17720 (1-3559)
QY 14 G1yleuValleuserAlaCyglInserThrProIleProProlyAsnAnpProGln--- 32
Db 585 GGCCGCTGCTGCGAGCTGCGAGTCCGTATCAGAAACACTTAAAGCTGATCCCTCATAT 644
QY 33 -----LeuAlaGlnIleArgThrGlnIleAlaIleSerLeuAsp 46
Db 645 AAACACAGACAGAAATCCCTTGAATGTATGCTATATGGACACACTTAAAGCTTAAAGCTT 704
QY 47 MetG1ylLeuSerAlaAlaGlnIleuSerAlaAlaLeuSerAlaAspArgGln 66
Db 705 GCTGGCAACAGTCAGAGGAATTCAAAATACATACAGTATCAAGATTGATTCACAC 764
QY 67 PheAlaProAlaTyraGthrLeuAlaLysValTyrglnIleAspGluAspAlaThrHis 86
Db 765 TATCGCGCTCTTATATTAACCTTGGCGCTGATAT-----TCGAAATGATGACAG 815
QY 87 GlnThrValAlaGlnArgLeuPheGlnLysAlaIleGluLeuAsnProLyAspMetGln 106
Db 816 TACGACATGCGCCCTTAATGCTATGAGAGAGGCTCAATTGACGCGCCCATGATGCTGAA 875
QY 107 SerTyrmcAspTyrglyPheTyLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
Db 876 GCTTATTCGAATATGCGGTGTTATATACAAAATCGNGGGGATTGGAAATCGCAATTGCT 935
QY 127 TyrPheAspLys-----ProSerArgAlaIleGlyTyrgluGlyArgVal 141
Db 936 TGTATATAGAGGTCTTTAGCTGTCTCCCAATTTTGAATTCGCAAGAACCAACATGACA 995
QY 142 ValAlaIleGluAsnMetAlaTyrlIeTyrtIleGlnTyrgluAlaIleLysSerPro 161
Db 996 ATTGCTCTGACTGATTA-----GCCACAAAGCTTAA 1028
QY 162 ThrIysAspAspTyraAsnAlaLysSerAlaLeuGluArgAlaLeuIle----- 178

```

```

Db 1029 TTGACAGAGACATCAACCAAGGTGTCATATTACAAAGAACCTTACTACATTCG 1088
QY 179 -----SerGlyThrGlnHisAspGluIleLysLysSerTy 190
Db 1089 CACTATGCCAGTCAATGATATATCTGGGTGTCCTATGCTGTAATGCTGAAA--TTT 1145
QY 191 AspLysLeuLeuSerAspTyrlLysLeu 199
Db 1146 GATATGCGCTATGCTATTTATGACTT 1172
RESULT 34
AE000798/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
Location/Qualifiers
1.12055
/organism="Methanobacterium thermoautotrophicus str.
Delta H"
/strain="Delta H"
/db_xref="taxon:187420"
/clone="MTH"
/note="synonym: Methanobacterium thermoautotrophicus str.
deltaH"
complement(118..375)
/gene="MTH58"
complement(118..375)
/note="Function Code:14.00 - Unknown, ; similar to,
pI:1N:S56131 AC:S56131, pI=0.99992, pI=15%"
/codon_start=1
/cran1_table=11
/product="unknown"
/protein_id="AAB84565.1"
/db_xref="GI:2621095"
/translation="MTGALIELGINSKORVPIDGVVPRDSYTLVETWPKGWLFGPY
TARTEIKYGRFHQTKTLERSDPIYIAPWLIILVLGATVM"
complement(172..953)
/gene="MTH59"
complement(172..953)
/complete="372..953"
/gene="MTH59"
/note="Function Code:14.00 - Unknown"
/codon_start=1
/cran1_table=11
/product="unknown"
/protein_id="AAB84566.1"
/db_xref="GI:2621096"
/translation="MKVRHLTALLMIIPIYGVVAPAVATGLWASPADFPYVNLKPGDSV

```

repeat_region	sp:LN:Y320_MYCEN, p(1)=0.81, pid=*
	/codon_start=1
	/transl_table=1
	/product="unknown"
	/protein_id="AAB84571.1"
	/db_xref="GI:2621101"
	/translation="MCHSTKFFFGVIVFVNVGPTVCSRLAIHTRPELPIIL TCLIGNMFLLIYLSKVVRRGEKGLVETLPTMSLIGVLLSMPT"
	complement(4365..5918)
	/note="1154 bp direct repeat includes part of MTH68 (O-linked GNAC transferase - paralog.fam. 17); 93% ID to interval 52093-50523"
	complement(4379..4545)
	/note="167 bp direct repeat includes part of MTH65 (unknown - paralog.fam. 20); 83% ID to interval 54315-54149"
	complement(4380..5140)
	/note="761 bp direct repeat includes part of MTH66 (unknown - paralog.fam. 20); 83% ID to interval 46441-45658"
	complement(4380..4763)
	/gene="MTH65"
	complement(4380..4763)
	/gene="MTH65"
	/note="Function Code:14.00 - Unknown, ; similar to, pir:LN:A39318, p(1)=0.98, pid=*"
	/codon_start=1
	/transl_table=1
	/product="unknown"
	/protein_id="AAB84572.1"
	/db_xref="GI:2621102"
	/translation="MAVAATIIATVGLHNTDDIKEFIIEHOTCILEDVPHGLDIAT VAGTAVMGPLGFISVVAEGDILIRTRTHYLPQEWKYSIHRTWIVGYELRSYGE DHCHEYIIRKNPGRSREMDAAYI"
	complement(4671..4904)
	/note="234 bp direct repeat includes part of MTH66 (unknown - paralog.fam. 20); 98% ID to interval 50200-49967"
	complement(4733..4780)
	/note="48 bp direct repeat includes part of MTH65 (unknown - paralog.fam. 20); 90% ID to interval 43075-43028"
	complement(4786..5007)
	/gene="MTH66"
	complement(4786..5007)
	/gene="MTH66"
	/note="Function Code:14.00 - Unknown"
	/codon_start=1
	/transl_table=1
	/product="unknown"
	/protein_id="AAB84580.1"
	/db_xref="GI:2621110"
	/translation="MIVAFGACVYCHOOTEWARELGOAIISEKDSLIDGNGTVLE GLDSTKXAYIGLCNHLVPHPPSPFSIM"
	complement(4870..5140)
	/note="271 bp direct repeat includes part of MTH67 (unknown - paralog.fam. 20); 94% ID to interval 54942-54669"
	complement(5127..5309)
	/gene="MTH67"
	complement(5127..5309)
	/gene="MTH67"
	/note="Function Code:14.00 - Unknown; similar to sp:LN:Y941_METUA AC:Q57711, p(1)=0.00054, pid=91%"
	/codon_start=1
	/transl_table=1
	/product="unknown"
	/protein_id="AAB84581.1"

Alignment Scores:	Pred. No.:	0.605	Length:	12055
Score:		111.50	Matches:	46
Percent Similarity:		43.24%	Conservative:	30
Best Local Similarity:		24.86%	Mismatches:	60


```

gene
3108..3923
/gene="TM1767"
/notes="similar to SP:P54382 PID:1303916 GB:AL009126
percent identity: 66.67; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="methyleneetetrahydrofolate
dehydrogenase/methylenetetrahydrofolate cyclohydrolase"
/protein_id="AAD36831.1"
/db_xref="GI:4982346"
/translaction="MMIDCKTARSIEERKEVEKLGFPKQVSVACDDBSALSYL
KSORKAEXKGLIAFEIMNVSPEEIVSTLKGSDSVNGVVARPEPLGIDEKEILSS
VPEKDEGVGNPALGILLDEEIPFCTAAVAILERETNLGKRVKVMVRSVTVG
KPLALMLKKGRDATTVCSTRVNLBEITKNSDIYVAVGRAFLPKMVKKCAIYI
DGINVYDGKLGQDVPSVEIARVTPVPGVGVQVTTALLFHHVAAERQR"
3920..5104
/gene="TM1768"
3920..5104
/gene="TM1768"
/notes="similar to GB:I42023 SP:P43913 PID:1003678
PID:1222325 PID:1204647 percent identity: 66.67;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="exodeoxyribonuclease VII, large subunit"
/protein_id="AAD36832.1"
/db_xref="GI:4982347"
/translaction="MKDYTSVTEINEYIKDLIEGPVLTNVSVEIGISVPRKGI
FESVLENALKECVIFEGDMWGIKLOEGRALVEGYSVITPPTTRFICSNRYIDQ
AGMYQIKFETTKLLEGLSLRPKKTVPKPIGIIITSRDAALQDVIRYREKKA
PIRIYVHTSVQDSAREELIKALRKANEDLDLWIVRGGSGKEDLWVFNEDVIRE
ILRLRHVTVIGIHEIDRVADVADVSMTPGAAEVYIPDASEIHEDDSFEKLI
TSLSNRPMERLETLYPRLMIGRKELNFKIERVEKLEAKLRKKLMDFEHNO
KRLDSLRMLSESLNPLRPLRGRFVLYKKDKIYKESIDLKRGDVSLVFPDGTKKQV
IG"
5116..5328
/gene="TM1769"
5116..5328
/gene="TM1769"
/notes="similar to GB:D00694 SP:P22938 PID:216583 GB:U00096
PID:1773106 percent identity: 70.31; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="exodeoxyribonuclease, small subunit"
/protein_id="AAD36843.1"
/db_xref="GI:4982358"
/translaction="MMKELEIVNRLNEDLPLESIKLFGVEVELYRKCKELIQNR
LKIIVKKELEGEIDASGRDOENELR"
5291..7117
/gene="TM1770"
5291..7117
/gene="TM1770"
/notes="similar to GB:U00096 SP:P77488 PID:1773104
PID:1786622 percent identity: 65.49; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="1-deoxyxylulose-5-phosphate synthase"
/protein_id="AAD36833.1"
/db_xref="GI:4982348"
/translaction="MLLDEIKRMSYDELKLAIDIRKITEVVLKNGGHLASNLGTE
LTALIVPDPREDATIMDTHQAYTHKILITGRDPLFHTIRTRGGISGFTPRESPID
WRGTGAGTISAGLGEKAFELIGERHVVVITGCAALSGALBLANDIKULNKM
KIILNDGMSISFVGGALYHLSKLRSPILYLGKVKLVLEKTEIGFVESEMKY
RSLDKMTIGTINFESLGLKFGFPDHNTELEKVKRLRDYDSSVHVVTKKKG
PTAAERPTKTHSASPGKPKMLSYSELGHTLSRYAREBKIVALTAAADATGTSI
FOEHPRDFDLITGQTCVTFGAALGTHGKMPVVAIVSTFLGRAVDQIHDVALGNA

```

```

gene
7114..7443
/gene="TM1771"
7114..7443
/gene="TM1771"
/notes="similar to GB:AL009126 percent identity: 51.35;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36834.1"
/db_xref="GI:4982349"
/translaction="WRIMLENGELEITNAKLIYRPTLESYGVGLDSSPFERI
FGSDGIGIKLEIEDSLNDVYIEVEYGNIREVAKNIDVNAHKLTLAGCEKLNIT
VHVVGIR"
7554..9047
/gene="TM1772"
7554..9047
/gene="TM1772"
/notes="similar to GB:AL009126 percent identity: 60.53;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36835.1"

```

Alignment Scores:

	Pred. No.:	Score:	Length:	Matches:	Mismatches:	Indels:	Gaps:
US-10-069-544-2 (1-204) x AB001815 (1-15428)	0.935	111.00	15428	54	38	75	11
Best Local Similarity:	41.63%	24.43%	10.64%	55	55	55	11

```

QY 22 SerThrProIleProProlysaAsnProGlnLeuAlaGlnIleArgThrGlnIleAla 41
    |||||
DB 9519 TCTACGAAAAGCCTCAAGAGTCAACAAAGCTTCTCTTACACCTCCCAAGAGAG 9578
    |||||
QY 42 IleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaLeu 61
    |||||
DB 9579 ATCTCCTTCTGAAAGAGAGAGGCTCGAAGAGCATGGAAGATACGAGAAATCTC 9637
    |||||
QY 62 SerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSer 81
    |||||
DB 9638 GAGAAAGACTCCAAATTTTGTCAAGTCVACGAAAGCTGGAGTTATCTACATCAGCTC 9697
    |||||
QY 82 GluAspAlaThrHisGlnThrLysAlaGlnArgPheGlnLysAlaLysGlnLeuAsn 101
    |||||
DB 9698 CAA-----CGATTCAAGAGACAGAAAGTCTTCAAAAGGCTTGGAAATAGAA 9748
    |||||
QY 102 ProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeu 121
    |||||
DB 9749 CGAAAGATCATGTGGAGTAC---ATCTCTCTTACATCTCATATAAATCTCGAAATCTC 9805
    |||||
QY 122 SerGlyAlaLeu-----IleTyrPheAspLysProSer-----132
    |||||
DB 9806 TTGGAAGCTCTGAAATCTTAAGAGACTTTATGAAAAAACCAGATGATCCATGATA 9865
    |||||
QY 133 -----ArgAlaIleGly---TyrGlnGlyArgValAla 143
    |||||
DB 9866 GCCAAGAAATACGAGACTCTCTCTCAAGACCTCGGCTGTATAGAAAGCCCTCGAGGTC 9925
    |||||
QY 144 IleGluAspMetAlaTyrIleTyrThrGlnTyrGlnAlaLysSerProThrLys 163
    |||||
DB 9926 TTGGAGAT--GCTTACCGCAGCATTAAGAGAGAGATTTCTGAG-----9970
    |||||
QY 164 AspAspTyrAsn-----AsnAlaLysSerAla 172
    |||||

```

Db 9971 -----TACAACTACGGAAGATCTCTCACTTGTAGAGAGGAAGCCGATTCATC 10024
Qy 173 Leu-----GluArgAla-----LeuIleSerGlyThrGln 182
Db 10025 CTCCTCGAAATCTCTGCTGAGCTGAAAGATGAGCCGAGTTCATCTCACTCCGAGAA 10084
Qy 183 HAspGluIleIleYelYelSerTyrAspLysLeuLeuSerAspTyrIlySLeuLeuSerAap 202
Db 10085 AAGGATGCTGTATCCCTCTTACGAG-----GAATTCGAATGCGCTGAAAGAT 10132
Qy 203 Tyr 203
Db 10133 TAC 10135

RESULT 36
AE004798/c 17148 bp DNA linear BCT 30-AUG-2000
LOCUS Pseudomonas aeruginosa PA01, section 359 of 529 of the complete
DEFINITION genome.
ACCESSION AE004798 AE004091
VERSION AE004798.1 GI:9949963
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.

REFERENCE 1 (bases 1 to 17148)
AUTHORS Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,
Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,
Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y.,
Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.
TITLE Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
JOURNAL Nature 406 (6795), 959-964 (2000)
MEDLINE 20437337
PUBMED 10984043

REFERENCE 2 (bases 1 to 17148)
AUTHORS Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,
Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H.,
Hancock, R.E.W., Lory, S. and Olson, M.V.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE Location/Qualifiers

FEATURES
source 1. 17148
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
gene complement(91..552)
/gene="PA3791"
CDS complement(91..552)
/gene="PA3791"
/codon_start=1
/product="hypothetical protein"
/protein_id="AG07178.1"
/db_xref="GI:9949964"
/translation="MTTLKRGQIGMALFAMLMITFVGPLISQSPHMPHMGVGM
GHEGHEHNDHDMHAGCSDAORSVTHAGFMGEGEGNRLASVLEKCGCSLL
FHSPLTEADCAFLSVFPATAPAVAFLEAFVASFVFGATRAPFLIA"
924..2702
/gene="leuA"
924..2702
/note="PA3792"
CDS /gene="leuA"
/codon_start=1
/transl_table=11

gene
/product="2-isopropylmalate synthase"
/protein_id="AG07179.1"
/db_xref="GI:9949965"
/translation="MPGPATNMLSDASHLVTAPAMPRRPBNRORISPMMLKDPG
QKTRFSAINLPDRTPESKTTTEPDTRKSSDLRDNQSLIEPMDAKKRPFRTYOV
GIKQIEVAPPSASDIDFNFRVLEIGNIIPDVTIQTQAREDLITRFEELRAKK
AIVHVNATASFRRIIVNODKQGVADTATNAKLIKLAAEQEPITQMSFOYSPIFS
STELFSEVCNAIVDWQPTPDNKIILNLPATVCACTPNYVQDLVTLLMNYTGGH
VILSIHTNDGTGVAATELGIMAGDVEGGLPENGERTGVADLIRGFPAGQKEDAI
WEVPLPIDPDIGRDYEAVIRVNSQSKGITFLIEBPGISLERRQIESQVYQ
ETDRGLGEMTAQOITISLENEYLKATISPYLASHRLQSENGEAVDLVSFDGEQHW
RGIGKGPLEALVAALPVKAEIMDYHEHAIGANAKAAAYIEIRLEGORPLIGIDE
NITTSFPAISALNRAVTOAEAKAA"
complement(2773..3105)
/gene="PA3793"
CDS complement(2773..3105)
/gene="PA3793"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AG07180.1"
/db_xref="GI:9949966"
/translation="MTGTYLLDELLETADMLEIDGLHAMRELENNLLDQDLAED
QPFASIEDVVLAVESLDGRTREWRFSYNAWEAEQADGESWRLTTGGAVALRCLGA
VSASGEDE"
complement(3159..3614)
/gene="PA3794"
CDS complement(3159..3614)
/gene="PA3794"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AG07181.1"
/db_xref="GI:9949967"
/translation="MNSDTENFRPPARLDEPDATGSEPLVRLSAGIGTFITPLA
GAFLAANMLRRLGRAQEVKTEWLVGLCLFVLPVGLIPENIPISIGTFVADIFGMVY
YAKSAFGPALDSHKAAAGAFISNWRAGIGLLFMLVLSVAIPVMLVY"
3795..4745
/gene="PA3795"
CDS 3795..4745
/gene="PA3795"
/codon_start=1
/transl_table=11
/product="probable oxidoreductase"
/protein_id="AG07182.1"
/db_xref="GI:9949968"
/translation="MTTRKLGNSGLEIPALVFGNVFGMTADESTFRLLDALDAGL
NCIDTADYSRVVPQHGGSESTLIGKMLKRTGKDRVVIASIKVMDMGNGHKLGSAA
YIEOLAEISLRILTQTDVLDLYOSHDDPDHDTLEETLSTYGLGELIKKQVIGVANSYDA
RLLEAFROVSARINLPSYOSILOPEVNTLDPADYEETNLEPTVEELGIGVIVSYSLASGF
LSGKTRNQDTRNGRARGERKVKYTLNERGVAITAALEDAVEQNNANPTOVALLMLIARP
TVTAPIASATSLIEQDLIDLIANTHLKLDQALHRDSASAY"
complement(4833..5420)
/gene="PA3796"
CDS complement(4833..5420)
/gene="PA3796"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AG07183.1"
/db_xref="GI:9949969"
/translation="MGGLDSITLGSALGSGEGCGGARDPVALLIGLVMTMSRGGAA
OGCGGGLISLISLIGAGGGAASGELGVTLLGGAAGAQAQAPRVASLISG
LGLGLQTMQNGMLGALNSWITGGGPNQSVASADISRVGDDQLQDLASAGVSGSAA
HLSLSLPELNVKLLTPDQAPGQDGLISLARFS"
complement(5554..6348)
/gene="PA3797"
CDS complement(5554..6348)
/gene="PA3797"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

```

/protein_id="AA007184.1"
/db_xref="GI:9949970"
/translation="MRDITVLPDIDILVQSSLYVHDAQANREHFAALLSAGADIV
VLPWFITGFSASAEQAEPLGPTTHMLLEQARLCAVVGSLIYLAAGSHNNRL
WAPDGEMLHYDKRHLFRMAGEHEHSPGEQELFKGKRVRLICYDLRFVWSD
PHDIDLITYANMPAPRQHWNRLLPAPALINLCYVAVNRIGEDNALRYAGDSQVL
DFQDSLFENADADGFFVRDLDAALAAVGRFPAVYADAFELHP"
complement(6336..7484)
/gene="PA3798"
complement(6336..7484)
/gene="PA3798"
/codon_start=1
/product="probable aminotransferase"
/protein_id="AA007185.1"
/db_xref="GI:9949971"
/translation="MIQSKLPVGTITFTMSQIAEACGAINISQGFDPDGPPELRE
AVGRHIAAGHNOYAPMTGLPALREQVAKKLEQFYGRVSDSEVTITPGATEAFSAI
QALVHPDEVVLDPSYDSEYVQLAGSCVHPALPDESIDMORMAAIGPRRL
VIINTPHNSGALISRELDRIALIDRIYILSDVEHVLVDDGVAHVSJHDEL
YPRSFVVSFGKTYHVTGKTYVAPPAISAEIRKHYQVNFQCTPLDMALADTMA
GHPHRLAQLPQFYQAKDLFCDDLDSRFRTAPGYTFCQVDISALRPDLDVANAE
MLTREGVAAIPSVFYEKAPDAMRLVRFYAKREETLRQAEKLCAI"
complement(7792..9273)
/gene="PA3799"
complement(7792..9273)
/gene="PA3799"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA007186.1"
/db_xref="GI:9949972"
/translation="MVPVIALNGRPVNGKSTLFNRLTKSPDAIVAYAGLTDRQGE
ARWGRYIYIDTGGISGDEGIDAKAEBSDIAEADAVLVSRKMTADQMTI
AEHRKRNKRSFLIANKVDITIDPLAEPSPGLGADPIAAHGRGNIHMLQELG
IIPKDNAEEGESEPAEEVAEGBEPRIIPGSEKGIKAIIGRPVNGKSTLVNML
GEERVIYDQAGTTRDSIYIPERNEKYTLIDTAGRRRGIKEANEKSPVYKTLQ
IODANVYIEMDAREGVHEHDLNLGFLVETGRALVIALNKKMEAEAREKYTELE
RLLFVDFADIFHISALHGTGVGHVXSVQSPRSATVTPSTLSILEDANQVHQP
PMNNGRIKIRYAHLGANPPLVYHNOVDANPKATRIETKTYRVLVLTGTPIRI
EYKGENPYEGKNSLTARQVNNKRRILMSHKKAEKKKDKRR"
complement(9351..10493)
/gene="PA3800"
complement(9351..10493)
/gene="PA3800"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA007187.1"
/db_xref="GI:9949973"
/translation="MYOMKHAALLALALAVGSSSKKEIPAEELTDRKEVYLSKO
WRSVGDGCGDLYNLBPAVDSTIYAASBEGVMAIORETDVLMKKDLERVSQGV
GVGYGLVGTGLDVTALDEATGKKKTRVNSVLSAPATNGVAVVQTDQDKLIG
LDAASDQRMVTESTVPLTIRGTGAPLAGNMAAGLAKGVAAVDVORGPIWEOR
VALPQGRSELDVVDIDGLILSGDLYVVSXGAAALDVNSGRLLMORASYYGV
AEGFGNIYVSOAGSVEGLDSGASLMMNDALAROLSPAVFSNVVVQDLEGVH
LISQVGRFVGRVSDGVRVPLVVGSMVYFNGKLVATTIR"
complement(10495..11139)
/gene="PA3801"
complement(10495..11139)
/gene="PA3801"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA007188.1"
/db_xref="GI:9949974"
/translation="MTRTEDEDDIAQIKDMQNRKPLVTGGLALAVFGQAWQKY
QTNQAGASIVYQOLIETALNPAGQDPKPAVLAELAGLSESGTTHAYVQGLFPAKY
AVDSGLDAVAELKAVMOKPADATIGELARORLAVLAAQKVERGKILGDDAKA
YSTRELRGDLVLQKRDDEAPAAVEKAKQSLPEGADALQIKLIDLSKEBA"

```

Alignment Scores:
 1.07 Length: 17148

```

Score: 111.00
Percent Similarity: 37.798
Best Local Similarity: 24.424
Query Match: 10.644
DB: 1
Gaps: 7

US-10-069-544-2 (1-204) x AE004798 (1-17148)

QY 1 MetLysIleArgValIleTyrProMetValMetAlaMetGlyLeu-----ValLeu 17
DB 15384 ATGACTGTACGCGCCGCGCTGTGTCTTGTGGCAGTTGGGCTGACAGCTGCGGAGC 15325

QY 18 SerAlaCysGlnSerThrProIleProPolysAsnAsnProGlnLeuAlaGlnIleArg 37
DB 15324 AGCGAGATCAGATCCCTCTGAAAACCGAAGGCGCGTACAGAGCGCGCAAGCTTAT 15265

QY 38 ThrGlnIleAlaIleSerLeuLeuAspMetGlyIleLeuAspGlnAlaIleGlnIleu 57
DB 15264 ATCCAGCTTGCGCTCGCTATCTCCAGCGGGGAAACACCGACAGCCAGGTCGCGTG 15205

QY 58 AspAlaIleLeuSerAlaAspArgIleAlaProAlaTyrArgThrLeuAlaIleVal 77
DB 15204 CGCAAGCGCCCTGAGATCGACCTTCCAGCGCGCAGCGCAGCGCGCTGCGGTGTA 15145

QY 78 TyrGlnAlaSer-----GlnAspAlaThr 81
DB 15144 TTCAGACCGAGATGAGCCCAAGCTGCGCAGAGAGTACCGAAGCGCTCGCCAGC 15085

QY 82 -----GlnAspAlaThr 85
DB 15084 GATPAGCCGATGCCCGGCTACTGAACAACGAGCGGTTCCTCTAGAGCAGAAAGCGC 15025

QY 86 HisGlnThrLysAlaGlnArgLeuPheGlnIleGlu-----LeuAsnProLysAsp 104
DB 15024 TACGAGAGAGCTTACAGAGGTCTCTGAGAGGACAGACACCTCTATCCAGCGC 14965

QY 105 MetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAla 124
DB 14964 TCGCGGCTTTCGAAACCTCGCGCTGTTGCTGCGCAATGAAGAAACCGCGCAGGC 14905

QY 125 LeuIleTyrPheAspLysPheSerArgAlaIleGlyTyrGlnGlyArgValAlaIle 144
DB 14904 AAGGAGTACTTCGAAATAATCCCTGCTG---CTGAACCGTAACACCGACGCGCGCTG 14848

QY 145 GluAsnMetAlaTyrIleTyrTyrHisGlnTyrGlnAlaAlaLysSerProThrLysAsp 164
DB 14847 GAG---ATGGCCGACCTGCTTACAGAGAGCGCGAGTACGTG-----CCGGCTGCGCAG 14797

QY 165 AspTyrAsn-----AsnAlaLysSerAlaLeu 173
DB 14796 TACTATGATCTCTTTCGCCAGGCGCGCGGAGGAGCGCGCGAGTCTGTTG 14746

RESULT 37
SME591785/c 286550 bp DNA linear BCT 05-JUL-2002
LOCUS Sinorhizobium meliloti 1021 complete chromosome; segment 4/12.
DEFINITION AL591785 AL591688
ACCESSION AL591785.1 GI:15073719
VERSION
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE
1 (bases 1 to 286550)
Capela,D., Bailly-Hubler,F., Gouzy,D., Bothe,G., Ampe,F., Batut,J.,
Boisset,P., Becker,A., Boutry,M., Cadieu,B., Dreano,S., Gloux,S.,
Godie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Maany,D.,
Pohl,T., Portetle,D., Puhler,A., Purnelle,B., Rampeger,U.,
Renard,C., Thebaud,P., Vandewol,M., Weidner,S. and Galibert,F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
JOURNAL MEDLINE 21396507

```

PUBMED 11481430
REFERENCE 2 (bases 1 to 286550)
AUTHORS Gouzy J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELLIO
EU Consortium
COMMENT
MELLIO EU Consortium:
Laboratoire de Biologie Moléculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Génétique et Développement UMR6061-CNRS,
Faculté de Médecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Université Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1148 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculté
des Sciences Agronomiques de Gembloux, Avenue Machechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:jerome.gouzy@oulouise.inra.fr
http://sequence.toulouse.inra.fr/melliot1.html.
Location/Qualifiers
1. 286550
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
70. 813
/gene="gac11 OR SMC00916"
70. 813
/gene="gac11 OR SMC00916"
/fc_number="2.5.1.18"
/function="miscellaneous"
/note="Product confidence : putative"
Gene name confidence : putative
predicted by Codon usage
predicted by Homology
predicted by Framed
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein_id="CAC45361.1"
/db_xref="GI:15073720"
/db_xref="SPTREMBL:Q92RR5"
/translation="MADLSFPTTAWPAKNPDIIQLVSLPTPNGVKVSIATBELGLA
YEAHRISPTNOKSPESFLNPNRIIPAIIDNPGDGKPIGLFESGALITLYAEKTG
KLIPADARVETLCVMFOMGIGMPFGHFKFKADKVAANNYPYORRDEKRR
LGLVLEARLEGKQMLMGKRNTPSPTRPIPMKGRSSNVAARCSTITRASRNSWTCMI
AAWHALRRRARRSRKSRSRANSLPAGD"
588. 779
/gene="SMC00915"
588. 779
/gene="SMC00915"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical"
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC45362.1"
/db_xref="GI:15073721"
/db_xref="SPTREMBL:Q92KLS"
/translation="MADGSEYTIADIATYPMVEGARFKFYGAELVDYKSPFVNDWV
DRGLAPAAQKGEIPIRKE"
867. 2084
/gene="SMC00914"
867. 2084
/gene="SMC00914"
/function="small molecule metabolism"
/note="Product confidence : putative"
Gene name confidence : hypothetical

repeat_region
predicted by Codon usage
predicted by Homology
predicted by Framed
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAC45363.1"
/db_xref="GI:15073722"
/db_xref="SPTREMBL:Q92RR4"
/translation="MSGRLVYVGGQAPALVAKURLAKDMRPTVVAEASLPYRQ
PLSKVYLREMTLDRLTYRPEAWAEHIDIRLSTTVRVDRLAQVALSDSMILYE
TLAFYGTGTPRLPAVAGDLAGVVEIDFRDADLAEMQGRRLVVGCGYIGLEA
AAVARTSGLEAVIMADRIQORVASTAIVREIHRSHGVDIRRTVGLIGDNG
RTYAELEDGSIPIVDIVIGVANDALAHAEAGIETANGIVDSHGRTSPPTVAM
GDCAVLPMDCMRIRPESVONAVDSEANVAIAGCTDYPDKPMFMSPOYKQIAG
FELGDELTVROGQRQSSVSWYFROGKLIVADINDAKAYVTGKKLBAGTPIRL
LADPVDILALIV"
complement(2114. 2572)
/note="Sm-2 OR SMC04876
REPEAT SM-2
predicted by Homology"
/evidence=not experimental
complement(2585. 4222)
/gene="groEL1 OR SMC00913"
complement(2585. 4222)
/gene="groEL1 OR SMC00913"
/function="cell processes; chaperoning"
/note="predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="60 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45364.1"
/db_xref="GI:15073723"
/db_xref="SWISS-PROT:P35469"
/translation="MAAEKVRGRRARSKMLRGVILDAVVTLGPKRNVVINDKSF
GAPRTTKGVSAKEIELEDKFEENNGAQMVEVASKTNDIADGTTTATLVAQIVRE
GAKVAAGNPMDLRGIDILAEVYKLLAKAKINSDVAQVGTISANEKQIDG
DIAEMQVNGNEGVITVEARTAELELVSGMDRQSLSPYFVTNEKRVADIEDA
FLLHEKRLKSLNLAOMLVLEAVVQVSEGLLIABEVSEBALATLVNKLRGKLIAAV
KAPGDSRKNMLEDIALILGCTGVSEDLGLIABEVSEBALATLVNKLRGKLIAAV
GAGQSDLEGVAQIKQIERTSYDEBKIOERLAKIAGVAVIRVGAATVEVKEK
KDRIDANATRAAVQESIVRGCGALIRSSKITYKSENDQDQGNVIVRALOSP
KQIVENADDEASIVGKILERTDDFGYNAQTGYDMIMGIIIDPVAVRTALQDAA
SVASLLITTEMIATLPRKDAVAMPGCMGCMMDM"
complement(4298. 4594)
/gene="groEL1 OR SMC00912"
complement(4298. 4594)
/gene="groEL1 OR SMC00912"
/function="cell processes; chaperoning"
/note="predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="10 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45365.1"
/db_xref="GI:15073724"
/db_xref="SWISS-PROT:P35473"
/translation="WASTNRPDLHRVVRVRESSEKTKGIIIPDTAKEROGEIV
AVGSGARDESGVPLDVAGRIILFGWSTGEVKNEDDLIMKEADIMVGIG"
4863. 5690
/gene="SMC00911"
4863. 5690
/gene="SMC00911"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical"
Gene name confidence : hypothetical
predicted by Codon usage

/protein_id="BAB75989.1"
/db_xref="GI:17133426"
/translation="MTIAGRAPSRGMEFDVLDMLKDRDFEVMGSGILLPFCALAT
GGMLTGTFTVSTHGLASSYLSLEGANFLTVVSSPADMGHSLILMGPEAQGLTR
WQOLGMLPFWALHGAFCGLIGMLROFELARIVGRPRNALFSPRIAVFVSPLTR
LQSSWFRPSPSTGVAIFRFLPLQGFHMTVMGAVGLGALLCAHGAIVE
NTLFEDGGANTFRAPNFPQSEETSMWTANFPWQIPIGIAFSNKRMLHFFVLFPVPT
GLMSAVGIGIALNLRAVDVFSOEIRAEDEPETFYKILLNEGIRAWMAPODOP
HEKFPPEEVLNRGNAL"
2270..3649
/gene="pabc"
/db_xref="GI:17133427"
/gene="pabc"
/note="ORF ID: a114291"
/codon_start=1
/transl_table=11
/product="photosystem II CP43 protein"
/protein_id="BAB75990.1"
/db_xref="GI:17133427"
/translation="WTLISNTLGGGRDQSSGFAMWGNARLLINSGKLLGAVAHAG
LHVAGAVTLFEVNAHFIEKPYVEQGLLPHVATLKGVGGEVIDTFPYFVGV
LHLISSAVFGGIYHVAVGPETLEYSFPGYDMKKKMTNIIIGFHLIIGGALL
LVKAMFPGGYVDTHAPGSGDYRVITNPILNPAVIFGYILKSPFGEGEIVYNNMED
VVGHWLWLCISGIMWHIFKPGEMARAFIWSGEALYSTSLGALSLMGFTASCW
WYNTVYSEPFPGPTGPBASQAALFLVRDRLQANVGSAGPTGLCYLMRSCE
IIFGSETRMFDFQCGWLEPLRGNLIDDKIKNDIQWQARRAAEYMTHAFLGSLNS
VGVVATEINSFVYSPRAWLATSHPVLAGFFFLIGHLMHGRARAAAGPEKGIIDRENE
PAMAGDLD"
complement(3926..4429)
/gene="a114292"
complement(3926..4429)
/gene="a114292"
/note="ORF ID: a114292"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB75991.1"
/db_xref="GI:17133428"
/translation="MRKYLGRATASLALINGYTIIPSOAIOLOGDTYFVQPRVLV
EVTYNNVNWGATYFTVYNLPENASEPLOGLTINOGCVNIRPDLKNSPFGCTR
SHRGORIGLKDADNRQTRTVSLIFDPVPSFGKTTITGLKPNQPTISGVYIFGYTAF
PAGEKNS"
4640..5131
/gene="a1r4293"
4640..5131
/gene="a1r4293"
/note="ORF ID: a1r4293"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB75992.1"
/db_xref="GI:17133429"
/translation="MNAISKVEFKRPSKQTALIPALGFMLASLVLDWVIMPSLYSG
KMLQGGFTTAGAIFWIEFNRLBLSAAVLTGILWKKTTHVMNNMTITGMLLAIA
LNTYFTFPQMSAIGVNLNLFATTESTIPATMNLHGXYFLIEVVKLVGALFNMCWQ
QOS"
complement(5294..6820)
/gene="a114294"
complement(5294..6820)
/gene="a114294"
/note="ORF ID: a114294"
probable N-acetylmutamoyl-L-alanine amidase"
/codon_start=1
/transl_table=11
/protein_id="BAB75993.1"
/db_xref="GI:17133430"
/translation="MDTISYSHASIYEASVEYIEPIKIVNLKFMQPKFSSNATIRL
LSVALSISLVAAGALALOKVSTSGSVEAIOREIKRIGFNGPFWGPFASITSGV
IGFORANRLADGVGGGTORALQACRSATGASISRGRLGSRGAATSOLOONRLR
LVEFNGPNTGYGSGEYTOCAVIRPORANRIGADGIYGNQTAQIRMAATSGVSEFVL
SSGSTCAVTRLOQLRLQJGYSPNFTGPRGRTTRDAVAFQKAGLPITGVVNOQTH
NALGVAQAQPNRPGSLTPQVRDQQLRLQJGYFNGPNTGAMTRDAVRFQDRRL
TADGIADVOILOAVSRWEDRYANQPNRNLTLTGDRGNVAVAOQLTQJGFTGSLD

gene
CDS
complement(7435..7917)
/gene="a114295"
complement(7435..7917)
/gene="a114295"
/note="ORF ID: a114295"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB75994.1"
/db_xref="GI:17133431"
/translation="MFSALEKFTIIGLILAIALLGSGVAVGLYFNNRPAIPETKMYA
NDTPALKDPSKKKIEBISTPOAKPEKNPPSPSTEKELPGGAYKRTWPGV
SLRAPTOEARVGGAGFNKKIIVYLENPDLMKKVRVSSKIBGWTGVNIERYDEE
"
complement(8048..9559)
/gene="a114296"
complement(8048..9559)
/gene="a114296"
/note="ORF ID: a114296"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB75995.1"
/db_xref="GI:17133432"
/translation="WNPFRBPKLLRLARVPLIYIPYVEERETAIRBEANQGNRPV
YTWDFVDYQGRNDAGFGRRNPLODALEFLKIPASAVAYIIRRYHFLUDVAIARK
LRNLARLKSQPKNTVLSPRIAPDDLEITLVEFPLPNAPEIKIETIRLOATGN
SLSEKVTDLVRSQGLSMERIRRVLAIAITHGLOEDVDVLEBRQRTROTQIL
DFYPATEQISDYGIDNLKDWLRRGGSFSEARQYGLPHRGILLVGIQCGKSLTA
KAIAHMHLPILRLDVGRLFGGLVGESESRTRONTQVABALPCLIMIDEIKASGL
KSGDAGTTSRVGTFETWLAKTSFVAVVATANNIOALPEMLKSGPDEIFPVGLP
SOEERKALFNVHLSLRPHNINKNDIEBLAETPSPGSAEIEQTILEAMHICFSQNRD
FTTDVLEASQIIPLARVAEQIOELQEMAAAGARLASKQSPUSDTFGKLK"
complement(9720..10376)
/gene="a114297"
complement(9720..10376)
/gene="a114297"
/note="ORF ID: a114297"
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB75996.1"
/db_xref="GI:17133433"
/translation="WYQIWNFLKEEYTFKKKINYNINPKLTWQVIDPFAVE
ONFLKSQLGEVETLEGFINLVVDNGYLRABEIAMFLVTRKQIWNLPDLAKTPSLK
DQLLSARNPOQFTKVAADGYFVSVDLAMLTEVKSPELVSINNSVGEILTVSSYG
KIEIGYIWLADWDGIVPPFCHRVQGTFLSQYIDNPFLPDRCFPKSYFKQOLVMWSH
"
complement(10817..12286)
/gene="a114298"
complement(10817..12286)
/gene="a114298"
/note="ORF ID: a114298"
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB75997.1"
/db_xref="GI:17133434"
/translation="NNQLFSIPALCLPAPETIALLQRTTIAMPKMFIRPQCKELLY
PADISIDLTSEORYRANLISVARTNIOSSQTFVIFAMARCELCQVADKTPNLV
SOLTITWPKAPEAIIQKQENETIYLYRTHLPQCKTVADNIOKRFKGVSLPINAG
SEDKPILKDSIFERRKQLEKTEPPQHELDLOSLFAPLAISNGAKLDDDIOTFL
GWKNTKQIQTLLNTDLAMINDIAKGDRIIEDEKSNYQAGTDFNITROSLEPFGFK
VEYAPAGAGGIDSLIECSQPYLVCSCKAGKSIPIBDAVEBELRIGRHKENYLOAVKL
IIEPQPKNLKDSLEISIKISINMTLOKVELVRYGPAIINLIELRYLPGQIDY
KLINEIDVEOAIKRSHIOLVNNYLTNSGSGEAGALHGAAYASHPPOPLKTEEM
HEIILESSPLTGYGRVKGDMDRDRYIYLRDLTISKK"
complement(12297..14201)
/gene="a114299"
complement(12297..14201)

Alignment Scores:

Pred. No.: 72.7 Length: 348550
 Score: 109.50 Matches: 54
 Percent Similarity: 38.72% Conservative: 37
 Best Local Similarity: 22.98% Mismatches: 81
 Query Match: 10.50% Indels: 63
 DB: 1 Gaps: 9

US-10-069-544-2 (1-204) x AP003596 (1-348550)

QY 13 MetGlyLeuValLeuSerAlaCysGlnSerThrProIleProPro----- 27
 Db 110335 CTGCGTTTACCACTGCCA-----CAACCTGCACAAATTACACAGCCGCCGACGATTT 110282
 QY 28 -----LysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSer 43
 Db 110281 TATCGCGCTACTCAAAAGTACCCCAATTTTGTTCCTTATTTGGCGGCTCA 110222
 QY 44 LeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnIleuAspAlaAlaLeuSerAla 63
 Db 110221 TTAATGGAAGGCAATTAATTACACAAACCAATGACTATTTCGAACGTCATTGGAATTA 110162
 QY 64 AspArgGlnIleAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAsp 83
 Db 110161 GAACCAAGTTGGGTTTGCACATTAATTAAGTGGCTTACTCAGACGACGACCAAAAT 110102
 QY 84 -----AlaThrHisGln----- 87
 Db 110101 TGGGAAGAGCGATCGCTCATTTCAAAAAGCAGTAGAGTTGAGTAAATAATGACCGGAA 110042
 QY 88 -----ThrLysAlaGlnArg 92
 Db 110041 CCTCATATTATTCTAGGATTTCTTATTATCAACAGCTAAATTAATGAAGCTAAATAAT 109982
 QY 93 LeuPheGluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112
 Db 109981 GCCTTCAATCAGAGCAATTAATCAATCAACCGCATTTCCGACGTTATTATCTGGGCG 109922
 QY 113 PheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLys----- 130
 Db 109921 GTTGTCTTATTATTAATCAAGGCCCAACCCAGAAAGCATTAAACAGCTTTAGAAAGTCAGCC 109862
 QY 131 -----ProSerArgAlaIleGlyTyrGlnGlyArgValAlaIleGluAsnMet 147
 Db 109861 GAAGCTAATCTTAATTAATCCCAATGCTTATTACGCT-----GACAGATTA 109817
 QY 148 AlaTyrIleTyrTyrHisGlnTyr-----GluAlaAlaLysSerProThrLysAspTyr 166
 Db 109816 GTTTTACCAATTAATCAATTAATGCGAGACGACAAAGTA-----TTT 109772
 QY 167 AsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGlyThrGlnHisAspGluIle 186
 Db 109771 AATCAGCCCAAGAT-----TTATATATGTAACCAAGGAATCCCAATG 109727
 QY 187 LysLeuSerTyrAspLysLeuLeuSerAspTyrLysLeuLeuSer 201
 Db 109726 GCAAAATAATCCGCAACATTTGTTACAAAGTACAAATTAAT 109682
 RESULT 39 SME591792 333800 bp DNA linear BCT 05-JUL-2002
 LOCUS Sinorhizobium meliloti 1021 complete chromosome; segment 11/12.
 ACCESSION AL591792 AL591688
 VERSION AL591792.1 GI:15075850
 KEYWORDS
 SOURCE Sinorhizobium meliloti.
 ORGANISM Sinorhizobium meliloti.
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Rhizobiaceae; Sinorhizobium.
 REFERENCE
 1 (bases 1 to 333800)
 Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J.,
 Boistard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S.,

FEATURES

source

gene

CDS

gene
CDS

Godt, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masny, D.,
 Pohl, T., Portetelle, D., Puehler, A., Purnelle, B., Ramsberger, U.,
 Renard, C., Thebaud, P., Vandenbol, M., Weidner, S., and Gallibert, F.
 Analysis of the chromosome sequence of the legume symbiont
 Sinorhizobium meliloti strain 1021
 Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
 MEDLINE
 JOURNAL 21396507
 PUBMED 11481430
 2 (bases 1 to 333800)
 REFERENCE
 Gouzy, J.
 Direct Submission
 Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
 EU Consortium
 COMMENT
 MELILO EU Consortium:
 Laboratoire de Biologie Moléculaire des Relations
 Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Caecanet,
 France, Laboratoire de Génétique et Développement UMR5061-CNRS,
 Faculté de Médecine, 2 avenue du pr. Leon Bernard, F-35043 Rennes,
 France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
 Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
 D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
 Université Catholique de Louvain, Place Croix du Sud 2, Bte 20,
 B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculté
 des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
 B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@univ-louvain.fr
 http://sequence.toulouse.inra.fr/meliloti.html.
 Location/Qualifiers
 1..333800
 /organism="Sinorhizobium meliloti"
 /strain="1021"
 /db_xref="taxon:382"
 complement(233..1114)
 /gene="SMC04043"
 complement(233..1114)
 /gene="SMC04043"
 /function="miscellaneous; hypothetical/global homology"
 /note="Product confidence : hypothetical
 Gene name confidence : hypothetical
 predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 evidence=not experimental
 /transl_table=11
 /product="CONSERVED HYPOTHETICAL PROTEIN"
 /protein_id="CAC47405.1"
 /db_xref="GI:15075851"
 /db_xref="SPTREMBL:Q92M30"
 /translation="MGEAVWEFPEFVLEPASQRIPEVNSPGRHYPQFLLDQSRD
 PHSIRSDHFDVLEFRSATFLGALRAHPRAVDVNRREYELDPMPDGLPFA
 NISSMRVAGGIGTVPRIVAENMEIYRGRRPPEQALERTETTKYPVATLRKLARTHV
 EFGMSILIDCHSMRPNVHLRSGSRHPDIIIDGRYTSAAELSRAYVLLBDLGAVA
 RKKPYAGGITEHVGRTPRGLHALQIEINRGLYDEATLIKPGPALEADATITAA
 LARHVEDFGAIVLPAAE"
 1305..1667
 /gene="SMC04044"
 1305..1667
 /gene="SMC04044"
 /function="miscellaneous; not classified regulator"
 /note="Product confidence : putative
 Gene name confidence : hypothetical
 predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 evidence=not experimental
 /transl_table=11
 /product="PUTATIVE 2-COMPONENT RECEIVER DOMAIN PROTEIN"
 /protein_id="CAC47406.1"
 /db_xref="GI:15075852"
 /db_xref="SPTREMBL:Q92M29"
 /translation="WTAKTILAEADNDNRRLVYALERAGYKVLSDYNGASAYDLRL
 EPPSILLTDIWMPEMDGIELARRATLELDPDKVMEITGFAAVAINPDSKAPDAVL

gene
CDS
1184. .1534
/gene="mlr2105"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB49318.1"
/db_xref="GI:14022710"
/translation="MDRPYLALAEFFSPKPGPYVLMALVALCALLTSATLCTGIYTSI
AALCTGVYLQNYRDTAAQAKLTNELIYALDTARNVVGLEHASPCAIDAEIKVENVN
ASTAPHEHTQDA"
/complement(1531. .2097)
/gene="ml12106"
complement(1531. .2097)
/gene="ml12106"
/note="hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB49319.1"
/db_xref="GI:14022711"
/translation="MIDCDARVARSIRFLSCGRAPKSPFNGSONAREEMPRRRCG
GNIULTENRHSCKGKPPDESIEMKETRNMAQAVAHATGTPMAFALCTASVLTAA
GVPVHYSTWQVLTNTGTTITFLMPLIQLNTQNRDGAIIQVKDELIRSLGKNDP
GIEHITREBEVEFRPACPAKATKASKG"
complement(2346. .2585)
/gene="ms12108"
complement(2346. .2585)
/gene="ms12108"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB49320.1"
/db_xref="GI:14022712"
/translation="MVSWFGPAWQSVFLGPEDLMAQCRIPOVRADANWERSAVSE
LAASVILAFQGFCTDEALTLAEIRSQDRRLMQRG"
complement(2682. .3116)
/gene="ml12109"
complement(2682. .3116)
/gene="ml12109"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB49321.1"
/db_xref="GI:14022713"
/translation="MTAFARAPRSPPLTIRKYSTILGMFDGLAGTTMKRRORS
SPFLYESHSPRRPSTAYGRILVANSQAPLAIAETEMAEHSDKVTALDCEILRGAF
RKSVEHRAVEREMHARVLAIRELDEIDPILMIYVK"
3560. .4291
/gene="mlr2110"
3560. .4291
/gene="mlr2110"
3560. .4291
/gene="mlr2110"
/note="hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB49322.1"
/db_xref="GI:14022714"
/translation="MSDPTMGSPQYRNELLIREMPADDDLALAPHMQRCLPYRMMLV
TPNPLIAVVFIEQIGSVASVASTASGHDAEIGLFGFGMTGSAIWMGDRAAHACFVQ
LEGAGATIDAPAFMAALASPTRLFLHFEVNTLHTQCTALVNAIKLEIRLARML
LMCDRDPGSLAMTHTEFLSITMLGVRRPGVTVALQLLEGBALIRSRGEIVIRBRL
LELANGYGOAEARVLAIRGEIMWG"
5065. .5808
/gene="mlr2113"
5065. .5808
/gene="mlr2113"
5065. .5808
/note="hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB49323.1"
/db_xref="GI:14022715"
/translation="MPIRKVLANLRRLVSGHSAVAVCRLNMNRTQFERLYQKGSV
PNATALICTYRIDEDDLYRAPEVAPSHKALLPIHQTLYNMVSGAPAVAGGY
FTFVTVDRDDLLVRSVTFVRARELVTFRVTRMRMGHGGGARVLGIMGVYVALISL
NMVYPAQINRRTREPSVMTVMQAPSPDPVYKRAFLITGSGPACVSVTRQVGPIS

```

gene      LNOVLGMCNRPVSLDPRDLQVLASLVREG"
CDS        5859..6530
           /gene="mir2114"
           5859..6530
           /gene="mir2114"
           /codon_start=1
           /product="probable transcriptional regulator"
           /protein_id="BAB49324.1"
           /db_xref="GI:14022716"
           /translation="MNSRDYSGKMDYTHALEBDITFGRAPGRLVDVLLAPRV
SHHTTROIYQLERLQIVTRENKGMWRKLSPEEVROIYEREMLQROALMILPLV
SEALJAQEEIHRISYGHVADAGYLRGHEANDRPHLTFMSAGMAYLVSSIEHYRLS
LVRANSLSADREKLDVSRQHHAMTEALKRDNWVLAQCVHLPSPKVFQYEIETA
PATG"
           complement(6597..7601)
           /gene="ml12116"
           complement(6597..7601)
           /gene="ml12116"
           /codon_start=1
           /product="probable zinc-binding oxidoreductase"
           /protein_id="BAB49325.1"
           /db_xref="GI:14022717"
           /translation="NMKAFVVDKYOQKGLRLAEMPEBLRDNVLYEITHAGVNLID
SLRDGEFEPILPYRPPFLLGHVAGIVVRAGSKARFKLGDEVARSDRGIGTFAQ
SIAMBADVALPKNLSMEBASIPLVGLTQWVLETAGLKKGVFOAGSGVGT
FALQAKHGATVATTSANANDLVKGGADVIDYKDDPEKTIQGVFVNSDQAK
TLEKSLHTKRGKGLISISGPPDPDPAGQKAMTKLVRLSLSGIRSKARRAGVG
SFLFMAGQDQGEITSRLESQFIRPVDIRIFPEKTEALAYVEAGRAKGVVKKR
"
gene      complement(7723..8538)
CDS        /gene="ml12118"
           complement(7723..8538)
           /gene="ml12118"
           /codon_start=1
           /translation="oxido-reductase"
           /protein_id="BAB49326.1"
           /db_xref="GI:14022718"
           /translation="MNRNSGRTAIVTGASSGIGRASAETLAAAGTIVETSKAGOS
SNGISMLVCDVTDASVASLVSIVLGTGRIDLNVNAGISGLGAEBSVSQGLF
DVNLFGVMKMTNAVLPMSMRLEGRIVNGISILGVAPYSNHYSAVKALGEGYESTL
DHEVAFNIRVSIPEAFVRTVFPDNGEPEDSVMKYDEARGLALLRDVPKADLP
EVAKVLRRAADRRHPLRRYTGAKAARQVSMRRAFPAGMFKILRKQRLPV"
           8432..8617
           /gene="mar2119"
           8432..8617
           /gene="mar2119"
           /note="unknown protein"
           /codon_start=1
           /translation="probable transcriptional regulator"
           /protein_id="BAB49327.1"
           /db_xref="GI:14022719"
           /translation="MPTVKPAPAKISALARPPEAPVTIAVPEPRFIVAPNVRG
LACDRILYDRNITLCS"
           8646..9233
           /gene="mlr2120"
           8646..9233
           /gene="mlr2120"
           /codon_start=1
           /product="probable transcriptional regulator"
           /protein_id="BAB49328.1"
           /db_xref="GI:14022720"
           /translation="NRDYGKRDASRSRIIMEVASHRFRGDGIAAGSLASIMSDAMTN
GAFYPHQSADLVRESMASALYETQSQOLQALAGGLELMAAMYLSPHRDHPQTGC
ASALLPEIARQPPETRSLYTEKLLTVHQLAKDLPRTRDPEGVALLGVFALLGALO

```

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:

81.1 Length: 346294
109.00 Matches: 53
38.72% Conservative: 38

```

Best Local Similarity: 22.55% Mismatches: 85
Query Match: 10.45% Indels: 60
DB: 1 Gaps: 8
US-10-069-544-2 (1-204) x AP002999 (1-346294)
QY 12 AlameTgLYLeuValLeuSerAlaCysGlnSerThrProIlePro-----ProIysAen 29
DB 66694 GGAATGCCATCTGCGCAAGGCCGACTACGACCGACCCATCACTACACCGAGCA 66635
QY 30 AenProGlnLeuValGlnIleArgThrGlnIleAlaIleSerLeu-----Aap 46
DB 66634 TCTTCTCGACTCGCAAGATCGGAACGATATTTCAACTCCGG-CTGGCCTGACCA 66576
QY 47 MetGlyLysLeuAspGlnAlaLysSerGlnLeuAspAlaAlaLeuSerAlaAspArgGln 66
DB 66575 AAGGCAATCTGAGCGCGCCATACGAGTTACACCCAGGCGCATGAGCTTCATCGCA 66516
QY 67 PheAlaProAlaIleArgThrLeuAlaLysValTyrGlnAlaSerGluAspAlaThrHis 86
DB 66515 CATGCCAAGCCTATCGCTGGCGCCGACCGCTGGTGAGCGGCGAT---ACGAC 66459
QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGluLeuAsnProLysAspMetGln 106
DB 66458 CAGGCCCTGTGCGAC-----TATACGAGGCGATCAGGCTGATCCGCGCGACCGCG 66405
QY 107 SerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
DB 66404 ACGTTCGCCACCGCGCGACGATCTGGAGCGCAACCGGACTACACCGCGCGCTCGCG 66345
QY 127 TyrPheAspLys-----ProSerArgAlaIleGlyTyrGlnLysArg--- 140
DB 66344 GATTACGATCAGCGCATGCGCTTCGCTCGCAACGACGCGGTGCTTACATGCGCTGCG 66285
QY 140 ----- 140
DB 66284 TGGATGTGTGCTGTAACATGACGACGCGCGCATGTGACTATGATCAAGCAACC 66225
QY 141 -----ValValAlaIleGluAsnMetCysAlaTyrIleTyrThrHisGln 154
DB 66224 GCCTTCGATCCGAACTATGCTGCTTACGACCAATCTGGCGCTGCGCTGCGTGG----- 66171
QY 155 TyrGlnAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlu 174
DB 66170 -----GATPAGGCGACCTTGACCGTGCATATCGGCTTCCAC 66132
QY 175 ArgAlaLeuIleSerGlyThrGlnHisAspGluIle----- 186
DB 66131 CAGGCCGTGATCGTGCATCCGAATATAGCGCGGCTTAATATGACAGGGGCTCGCGCG 66072
QY 187 -----LysLysSerTyrAspLysLeuLeuSerAspTyrLysLeu 199
DB 66071 ATGACCAAGAACAGTACGACCTTGCCATCGCCGACTCAATATG 66027

```

Search completed: July 4, 2003, 03:57:37
Job time : 2408 secs

THIS PAGE BLANK (USPTO)

XX 08-FEB-2001.
PD
XX
XX 27-JUL-2000; 2000MO-EP07281.
PE
XX
XX 30-JUL-1999; 99GB-0018040.
BR
XX
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA
XX
XX Thonnard J;
PI
XX
XX WPI; 2001-159868/16.
DR
DR P-PSDB; AAB60638.
XX
XX
PT New polypeptides and polynucleotides of *Moraxella catarrhalis*, useful
PT as vaccine for prevention, treatment of microbial diseases and in
PT diagnostic assays for detecting diseases associated with microbial
PT infections -

Claim 13; Page 66; 81pp; English.

The invention relates to the Moraxella catarrhalis strain ATCC43617 BABS121 protein (AAB6063b) and to DNA encoding it (AAFS9766b). The invention also relates to immunogenic fragments of the BABS121 protein, expression vectors and host cells comprising BABS121 nucleic acid, the recombinant production of BABS121, vaccine compositions comprising the BABS121 protein or nucleic acid, an antibody against BABS121, therapeutic compositions comprising the anti-BABS121 antibody, and a method of identifying a Moraxella catarrhalis infection via the detection of BABS121 proteins or antibodies. The vaccine compositions of the invention are useful as prophylactic or therapeutic agents against Moraxella catarrhalis infections in mammals, particularly humans. Moraxella catarrhalis is a Gram negative bacterium frequently isolated from the human upper respiratory tract, which is responsible for several pathological conditions. It is responsible for about 15% of otitis media cases in children (which can lead to temporary or permanent hearing loss). It also causes pneumonia in elderly people, and sinusitis, nosocomial infections and, less frequently, invasive diseases. BABS121 proteins or nucleotides may additionally be used in screening for novel antibacterial compounds, and in the diagnosis and staging of infections. The present sequence represents DNA encoding the Moraxella catarrhalis strain ATCC43617 BABS121 protein.

Sequence 615 BP; 200 A; 116 C; 132 G; 167 T; 0 other;

Alignment Scores:	
Pred. No.:	1.2e-118
Score:	1043.00
Percent Similarity:	100.00%
Best local Similarity:	100.00%
Query Match:	100.00%
DB:	22
Length:	615
Matches:	20
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-069-544-2 (1-204) X AAF59768 (1-615)

QY	1	MeLys1LeuValIysTRPProMetValMetAlaMetGlyLeuValLeuSerAlaCys	20
Db	1	ATGAAATCCGAGTAAAGTGGCTTAGTGAATAGGGGATGGGGCTGTTTAAAGTGCTGT	60
QY	21	GlnSerThrProIleProPolySaaManProGlnLeuAlaGlnIleArgThrGlnIle	40
Db	61	CAAGTACCCAAATACCGCTTAAATAATATCCGCATTTGGACCAATATCCGAATCCCAATT	120
QY	41	AlaIleSerIleLeuAspMetGlyIysLeuAspGlnAlaIysGlnGlnLeuAspAlaAla	60
Db	121	GCATCTCCGCTACTGACATCGGTGTAAGCTTGATCAAGCCAACAGCATTTGGATGCTGT	180
QY	61	LeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrIleuAlaIysValTyrGlnAla	80
Db	181	TTATCGGCTGATGGCAGTTTGACCTGCTATTCGACCTTGGCAGAAAGTITATCAAGCT	240
QY	81	SerGluAspAlaThrHisGlnThrIysValaGlnArgLeuPheGlnIysValaAlaIleGluLeu	1000

Dd		241	TCGGAGATGCGCACTCACCAACCAAGCTCAACGCTTTGTGGAAAAGCGATTGAAC	3008
Oy		101	AenP-polySAePM-eGlmseryT-metAspyTYrGLYPheTYrleuValGImerGYasp	1207
Dd		301	AATCCTAAAGAATATCCAAAGTAATATGTGATTATGTGATTTATTTGGTGACAAGGGGGAC	3609
Oy		121	LeuSERGIYAlaleuIleTYrPheaSPLysProSerArgAlallegLYTyrgluGIYArg	1407
Dd		361	TtGTcAGGTGCCTTAATTTATTTTATATTAACCCTAGCCGACCATCGCTTATGAGGCCGT	4209
Oy		141	ValVALAlallegLUsAmecALATyrIleTYrTYrhSGInTYrGUAlaalalySer	1607
Dd		421	GtGGTAGCGCATCGAAATAATATGGCATATATTTATTTATCATCAATATGACCTGCCAATCA	4809
Oy		161	PTrOTfRlySAeSPsPTyThASnaSuAlalySSerAlaleuGUlARGAlaleuIleSeRGY	1807
Dd		481	CCAA-CMAAAGAATGACTTATATATTAACCCCAAATCAGCACTTGACCGCTTAAATTTCAAGC	5409
Oy		181	ThrGINHsaAPGuIlleLYlseseryrAspLySLeuSerASPtyrLYslauLeu	2007
Dd		541	ACTCAACATGATGAATCAATCAAAAGCTTATGCAAAATTATTAAGTATTAATTAATTATTA	6009
Oy		201	SerAspTYrlys 204	
Dd		601	AGTGATTATAAA 612	
RESULT 2				
ID	AAF28536			
XX	AAF28536	standard; DNA; 33140 BP.		
XX	AAF28536;			
DT	04-APR-2001	(first entry)		
DE	Genomic fragment #23.			
KW	Genomic library; bacteria; human upper airway; otitis media; sinusitis;			
XX	bronchopulmonary; endocarditis; meningitis; ss.			
OS	Moraxella catarrhalis.			
PN	MO200078968-A2.			
PD	28-DEC-2000.			
PF	16-JUN-2000; 2000WO-US16649.			
PR	18-JUN-1999; 99US-0140121.			
PA	(INCYTE) INCYTE GENOMICS INC.			
PI	Lagace RE, Patterson C, Berg KL;			
DR	WPI; 2001-041427/05.			
PT	Genomic library for identifying diagnostic and therapeutic			
PT	compositions, and for identifying virulence factors, regulatory			
PT	elements and drug targets, comprises Moraxella catarrhalis nucleic			
PT	acids -			
XX				
XX	Claim 1; Page 191-199; 545pp; English.			
XX	The present invention relates to a Moraxella catarrhalis genomic library			
XX	comprising of a combination of 41 nucleic acid molecules (see			
CC	AAFP28514-AAFP28556).			
CC	The library has a number of uses described in the			
CC	specification e.g. is useful for identifying diagnostic and therapeutic			
CC	compositions. M. catarrhalis (Branhamella catarrhalis) is a large			
CC	aerobic, gram-negative diplococcus, normally found among the bacterial			
CC	flora of human upper airways. M. catarrhalis is known to cause acute,			
CC	localised infections such as otitis media, sinusitis and bronchopulmonary			
CC	infection and life-threatening, systemic diseases including endocarditis			
CC	and meningitis.			
XX				

SQ Sequence 33140 BP; 9472 A; 6496 C; 7384 G; 9787 T; 1 other;

Alignment Scores:

Pred. No.: 2 87e-116 Length: 33140
Score: 1043.00 Matches: 204
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-10-069-544-2 (1-204) x AAF28536 (1-33140)

QY 1 MetLysIleArgValIleSTPProMetValMetAlaMetGlyLeuValLeuSerAlaCys 20
DB 21062 ATGAAATCGGAGTAAAGTGGCTTAATGCGATGGCGGCTTGAATGCTGCT 21121
QY 21 GlnSerThrProIleProPheProLysAspGlnLeuAlaGlnIleArgThrGlnIle 40
DB 21122 CAAAGTACACCAATACCGCTTAAATATCCCAATGGCACAAATCCCAATCCCAAT 21181
QY 41 AlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAla 60
DB 21182 GCATCTCGCTACTGGACATGGGTAAGCTTGATCAAGCCAAACAGCATGGATGCTGCT 21241
QY 61 LeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAla 80
DB 21242 TTATCGGCTGATGGGAGCTTGGACCTGCTTACGACCTTGCAAAAGGTTTCAAGCT 21301
QY 81 SerGlnAspAlaThrIleGlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeu 100
DB 21302 TCGAGAGATCCCACTACCAACCAACCAAGCTCAACGCTTGTGAAAAGCGATTGAAC 21361
QY 101 AsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAsp 120
DB 21362 AATCCCAAAATATGCAAGTATATGATTAATGATTAATTTGTCGATGGGAGC 21421
QY 121 LeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGlnGlyArg 140
DB 21422 TTGTCAGGTCGCTGATTTATTTTGAATTAACCTAGCCGACCTCGTTATGAAGGCGCT 21481
QY 141 ValValAlaIleGlnLeuMetAlaTyrIleTyrThrIleGlnTyrGlnAlaLysSer 160
DB 21482 GTGGTGGCCATCCAAATATGACATATATTTATTCATCAATATGAAAGTCGCAATCA 21541
QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGly 180
DB 21542 CCAACAAAGATGCTATTAATACGCCAAATACGCACTTGAAGGCTGTTAAATTTCAAGC 21601
QY 181 ThrGlnIleAspGlnIleLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeu 200
DB 21602 ACTCAACATGATCAATCAAAAGCTTATGACAAATTTAAGTATTTAATTTATTA 21661
QY 201 SerAspTyrLys 204
DB 21662 AGTGATATATAA 21673
RESULT 3
AAV69302
ID AAV69302 standard; DNA; 4097 BP.
XX
AC AAV69302;
XX
DT 01-FEB-1999 (first entry)
XX
C. elegans OCT DNA.
XX
OCT; O-linked GlcNAc transferase; uridine; transferase; tumour;
KM diphospho-N-acetylglucosamine; polypeptide beta-N-acetylglucosaminyl;
KM predilection; type II diabetes; hyperglycaemia; Alzheimer's disease;
KM metastasis; diagnosis; inhibitor; treatment; diabetes mellitus; 86.
XX
Caenorhabditis elegans.
OS
XX

FR Key Location/Qualifiers
FT CDS 1..3456
FT /*tag= a
FT /product= "OCT"
FT /note= "O-linked GlcNAc transferase"
XX
XX W09844123-A2.
XX
XX 08-OCT-1998.
XX
XX 27-MAR-1998; 98WO-US06101.
XX
XX 31-MAR-1997; 97US-0042270.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Hanover JA, Lubas W;
XX
XX MPI; 1998-557118/47.
XX
XX P-PSDB; AAW82501.
XX
XX Protein exhibiting O-linked GlcNAc transferase activity, OCT -
PT useful, e.g. to assess predisposition to type II diabetes or
PT Alzheimer's or metastatic potential of tumours, and to identify
PT inhibitors
XX
PS Claim 4; Page 34-39; 56pp; English.
XX
XX This sequence encodes a novel *Caenorhabditis elegans* OCT, O-linked GlcNAc
CC transferase protein (also known as uridine diphospho-N-acetylglucosamine:
CC polypeptide beta-N-acetylglucosaminyl transferase). This protein is
CC useful to assess predisposition toward type II diabetes in patients
CC suspected of having hyperglycaemia that could evolve into this disease,
CC by assaying OCT activity in red blood cells. It can also be used to
CC assess predisposition toward Alzheimer's disease, to assess the
CC metastatic potential of tumours and to diagnose a tumour with metastatic
CC potential. OCT can also be used to identify OCT inhibitors, especially in
CC high-throughput assays, useful, e.g. in the treatment of diabetes
CC mellitus, tumour-derived diseases and Alzheimer's disease.
XX
SQ Sequence 4097 BP; 1252 A; 847 C; 776 G; 1221 T; 1 other;
Alignment Scores:
Pred. No.: 0.000113 Length: 4097
Score: 124.50 Matches: 44
Percent Similarity: 44.72% Conservative: 28
Best Local Similarity: 27.33% Mismatches: 64
Query Match: 11.94% Indels: 25
DB: 19 Gaps: 6
US-10-069-544-2 (1-204) x AAV69302 (1-4097)
QY 30 AsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuAspMetGlyLys 49
DB 769 AATCGATTTGATTTGTCAGAGTGAATCTTGGAACCTTAATAACCAATGGGAGAG 828
QY 50 LeuAspGlnAlaLysGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
DB 829 CTGGAAGAGCGAAGGCTGTTACTTGAAGCAATGGAATCAACCAACGATTCGCTGTC 888
QY 70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAsp-----AlaThrHis 86
DB 889 GCATGCTCAACCTTGATGATGATCATAGCAAGAGAAATTTGTTGGCAATTCAT 948
QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGln 106
DB 949 CAT-----TTGAGAAAGCTGTCACCTTGGATCCAAACTTCCTGCAC 990
QY 107 SerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
DB 991 GCTTATATAAATCTTGAAGATGTTGAAAGGCGCAGGATTTTCATAGAGCGGTTTCA 1050
QY 127 TyrPheAspLysProSerArgAlaIleGlyTyrGlnGlyArgValAlaIleGln--- 145

```

Db      1051 GCTTAT-----CTCCGAGCCTTGATCTCTGTGTATCATGCAGTTTTCATGG 1101
               |||||
Qy      146 AsmMelaTyrlleTyrlleGln-----Tyrglu 156
               |||||
Db      1102 AATTGGCAGTGTGTCTGTAACGAGGACTTATGATTTGGCATTGACATTACAA 1161
               |||||
Qy      157 ALAAla-----LysSerProThrLysAspAspTyrlleAsnAlaLysSerAlaLeu 173
               |||||
Db      1162 AAGGCTATCGACTTACACACCACTCTCCGTATGCTGTAATCTTGCAAAATGACTTA 1221
               |||||
Qy      174 Glu 174
               |||
Db      1222 AAA 1224

RESULT 4
AAT42063
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
KM expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO ) UNIV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
DR WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
XX Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
CC sequence at least 99% identical to (1). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
XX
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:
Pred. No.: 2.71 Length: 1830121
Score: 118.50 Matches: 39
Percent Similarity: 39.47% Conservative: 21
Best Local Similarity: 25.66% Mismatches: 89

```

```

Query Match: 11.36% Indels: 3
DB: 17 Gaps: 1
US-10-069-544-2 (1-204) x AAT42063 (1-1830121)
Qy      1 MetValIleArgValIleTyrlleProMetValMetAlaMetGlyLeuValIleuSerAlaCys 20
               |||||
Db      389428 ATGAAACAATAAGTAAACAACATGAGTGGGTATTTTCCCTTATTTTCTGCTGC 389487
               |||||
Qy      21 GlnSerThrProIleProIleProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIle 40
               |||||
Db      389488 GTTTCACAATCTGCATCTAGCTTAAATCATCAAACTGCCGCTTAAAGCCGCTGTGAACCTC 389547
               |||||
Qy      41 AlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnIleuAspAlaAla 60
               |||||
Db      389548 GCTTGAAGCTATCTTACGCAAAATTAATCTCAACTGCTTAAATCAATTTAGACAAGCA 389607
               |||||
Qy      61 LeuSerAlaAspArgGlnPheAlaProAlaTyrlleArgThrLeuAlaLysValTyrglnAla 80
               |||||
Db      389608 CTTCAACAGCATTAATAAATTACTATCTCGCATTCAGCACTTGACACATTATTATCAACA 389667
               |||||
Qy      81 SerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGluLeu 100
               |||||
Db      389668 CAAGGCGCAATAGAAAT-----GCCCTTCGGAATATGAATATACCCGTAACCTT 389718
               |||||
Qy      101 AsnProLysAspMetGlnSerTyrlleMetAspTyrlleGlyPheTyrlleValGlnMetGlyAsp 120
               |||||
Db      389719 AATCATTAACAAGCGCATGTGATTAATTTTGGACGTTTCTATGTCGTAAGAAAGAA 389778
               |||||
Qy      121 LeuSerGlyAlaLeuIleTyrllePheAspLysProSerArgAlaIleGlyTyrlleGluTyArg 140
               |||||
Db      389779 TTGAAACAAGCTCAGCAACAATTGAAATTAGCACTTAATTCGCCGAATATTATCATCAA 389838
               |||||
Qy      141 ValValAlaIleGlnAsnMetAlaTyrlleTyrlle 152
               |||||
Db      389839 GCAGATPACATTTGAAATAATCGCTTGTGCTTAT 389874
               |||||

RESULT 5
AAZ54213
ID AAZ54213 standard; DNA; 762 BP.
XX
AC AAZ54213;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 708 partial DNA sequence SEQ ID NO:2375.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalzo E, Scarbelli M;

```

PI Tettein H, Venter JC;
 XX WPI; 2000-062150/05.
 DR P-PSDB; AAY75451.
 XX
 PT Novel *Neisseria* polypeptides predicted to be useful antigens for
 XX vaccines and diagnostics -
 XX
 PS Claim 7; Page 1140; 1453pp; English.
 XX
 CC AA254515 to AA254536, AA254537 to AA254615, and AAY74253 to AAY75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria meningitidis* (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria meningitidis*, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SO Sequence 762 BP; 212 A; 222 C; 174 G; 154 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 8.1e-05 Length: 762
 Score: 117.50 Matches: 43
 Percent Similarity: 46.70% Conservative: 42
 Best Local Similarity: 23.63% Mismatches: 78
 Query Match: 11.27% Indels: 19
 DB: 21 Gaps: 5
 US-10-069-544-2 (1-204) x AA254213 (1-762)
 QY 4 Argvallylrtrprowetvalmetalametglyleuvalleuseralaacygsinsethr 23
 DB 22 CGAATCTCTTATTACTCGTTCTTGGCTTGGC-----GGTCCACACCTTCC 69
 QY 24 -----ProtleProPolylvaabaanaproglnleuvalaqlnllaargthrglnleu 41
 DB 70 TACCGCCCTCGCGCGGCAAAAAGCCATCAGGTTCCATATCAAAACCGCCTGCA 129
 QY 42 lleserleuueuapmetglylyleuapglalalaeslnglnleuapalalaieu 61
 DB 130 ATGGAATATATGCGCGCTCAGACTACCGTCAAGCGGCAAGTATGAAGACGCTCG 189
 QY 62 SeralaaparglnphealaProalaTyArgThreualaYsValTyrglnalaser 81
 DB 190 AATCGAGCCCTAAAACGAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249
 QY 82 GluapalaThrlsclnthrlYalaaglnargleupheglulYsalailegluleuam 101
 DB 250 AAA-----GTTAACGACAAAGCGCGCAAGTTCGCGCAAGCCCTCCATCAAA 300
 QY 102 ProlyaapmetglnsetryrmetlaPtyrGlyPhetyrleuval---GlnmetGlyasp 120
 DB 301 CCCGACAGTCCCAATCAACAACACTACGAGTGTCTTATGCGGACGCTCAACCGC 360
 QY 121 LeuserGlyalaLeuileTyrPheaplyPpseraYrgalailegTyrglnGlyarg 140
 DB 361 CCGGCAATCTATGCGATATTTGCAACAAAGCTGCGGCAACCCACCTCCGACCCCT 420
 QY 141 ValValaIailegluleuametalTyrlleTyrlsclnTyrglnalalaYsSer 160
 DB 421 TATATGCGCAACCTGAT-----AAAGCATATGCGAGC 453
 QY 161 ProthrlYsaaPpTyraaanaAlalaYsSeralaenglnuYrglnalaleuileseGly 180
 DB 454 GCAAAACAGGGGCAATTCGATTCGCGAAGCCATTTGAACGTTCCCTCGCGCCAG 513
 QY 181 ThrGln 182

DB 514 CCGCAG 519
 RESULT 6
 ID AAA81469/c
 ID AAA81469 standard; DNA; 60873 BP.
 AC AAA81469;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_17 SEQ ID NO.17.
 XX
 KM *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KM *Meningococcus B*; MenB; de.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN MO200022430-A2.
 PD 20-APR-2000.
 PF 08-OCT-1999; 99WO-US23573.
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 PA (CHIR) CHIRON CORP.
 PI Frazer CM, Hickey E, Peterson J, Tettein H, Venter JC,
 PI Maignan V, Galotci C, Mora M, Ratti G, Scarselli M, Scariato V,
 PI Rappnoli R, Pizzo M;
 DR WPI; 2000-318079/27.
 PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be
 PT used in the diagnosis and treatment of *N. meningitidis* infection and
 PT other *Neisseria* infections, for example, *N. gonorrhoea* -
 XX
 PS Claim 7; Page 428-446; 1760pp; English.
 XX
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA
 CC sequences. AAA81260 to AAA81303 and AAB25620 to AA25563 represent
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC *Neisseria meningitidis*. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseria*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SO Sequence 60873 BP; 14510 A; 14559 C; 16692 G; 15109 T; 3 other;
 XX
 Alignment Scores:
 Pred. No.: 0.0334 Length: 60873
 Score: 117.50 Matches: 43

Oy 121 leusergGAlaLeuilelyrPheapLySProSeRarGAlaileGlyrYrGlunGlyArx 140
 Db 127000 CCTGCGGCAATCTATGCGARATTTTCGACAAAGCTCTGCGGACCCGACCTACCGGACCCCT 1270539
 Oy 141 ValValAlaileGluAsnMerAlaTyrileYrTYrHleGlnTYrGluAlaAlaLySer 160
 Db 127060 TATATTTGCCAACCTGAAAT-----AAAGGCATATGCAGC 1270922
 Oy 161 ProTrHlyrASpAsPTyrASnASnAlaLySserAlaLeuGluArGAlaLeuileSerGly 180
 Db 127093 GCAAAACAGGCGCAATTCGATGTGGAGGAGGACCTATTTCAAACGTTCCCTCGCGCGCCAG 1271522
 Oy 181 ThrGln 182
 Db 127153 CCGCAG 127158
 RESULT 8
 AAA81490
 ID AAA81490 standard; DNA; 1437668 BP.
 AC AAA81490,
 DT 04-DEC-2000 (first entry)
 DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
 XX
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 XX Meningococcus B; Memb; ds.
 OS Neisseria meningitidis.
 XX
 XX WO200022430-A2.
 FN 20-APR-2000.
 XX
 XX 08-OCT-1999; 99WO-US23573.
 PE 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 PT Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
 PI Maignault V, Galeotti C, Mora M, Ratti G, Scarfelli M, Scarlato V,
 PI Rappuoli R, Pizza M.
 XX
 DR WPI, 2000-318079/27.
 XX
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 PS Claim 7; Page 866-1272; 1760pp; English.
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25653 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81234 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis Memb polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious

CC	Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC	Multivalent vaccines have also been tried but none have successfully
CC	overcome antigenic variability. The provision of further, complete
CC	sequences may provide an opportunity to identify secreted or surface
CC	exposed proteins that may be presumed targets for the immune system and
CC	which are not antigenically variable or at least more conserved than
CC	other more variable regions.
XX	
SQ	Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;
	Alignment Scores:
	Pred. NO.: 2.58 Length: 1437668
	Score: 117.50 Matches: 43
	Percent Similarity: 46.70% Conservative: 42
	Best Local Similarity: 23.63% Mismatches: 78
	Query Match: 11.27% Indels: 19
	DB: 21 Gaps: 5
US-10-069-544-2 (1-204) x ABA81490 (1-1437668)	
Oy	4 ArgVallysrTrpProwetValMelalameGlyLeuValleuseralacyGsinserThr 23
Db	1297050 CGAATCTCTTATATACCTGTTCTTCTGCTTGGCC-----GGTGACAGACTTCC 1297097
Oy	24 -----ProIleProPolysaenasnProGlnleuAlaGlnlleaGthrGlnlla 41
Db	1297098 TACCGCCCTTCGCGGACGAAAAAGCCAAATCAGGTTTCATATCAAAACCCAGTTGGCA 1297157
Oy	42 IleSerleuenuaPmetGlyLysleuaspGlnAlaLysGlnleuaspAlaAla 61
Db	1297158 ATGGAATATATGCGCGGTACGAGTACCCGACGACGCGCAATATTGAAGACGCCCTG 1297217
Oy	62 SerAlaaspArgGlnPheAlaProAlaTyrArgThrleuAlaLysValTyrGlnAlaSer 81
Db	1297218 AAATCGACCCCTAAAAACGAGCTTGCTGCTGGCTGCTGCGCAATCTATCATACCTG 1297277
Oy	82 GluAepAlaThrHisGlnThrLysAlaGlnArgleuPheGluYsaAlaIleGluLeuasn 101
Db	1297278 AAA-----GTTAACGACAGGCCACAGAAAGTTTCGCGCAAGCCCTCCATCAAA 1297328
Oy	102 ProLysaspMetGlnserTyrMeLaspTyrGlyPheTyrleuVal---GlnMetGlyAsp 120
Db	1297329 CCCGACGCTGCCGAATCAACAACACTACGATTGGTTCTATGCGGACGCTCAACCGC 1297388
Oy	121 LeuserGlyAlaLeuIleTyrPheaspLysProSerArgAlaIleGlyTyrGluGlyArg 140
Db	1297389 CTGCGCAATCTATGCGATTTTGCACAAAGCTCTGCGCACCCCACTACCGACCCCT 1297448
Oy	141 ValValAlaIleGluasnMetAlaTyrIleTyrThrHisGlnTyrGluAlaAlaLysSer 160
Db	1297449 TATATGCCCAACCTGAAT-----AAAGCATATGACGC 1297481
Oy	161 ProThrLysaspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGly 180
Db	1297482 GCAAAACAGGGCGCATTCGATTGGCGGAAGCCATTATGAACGTTCCCTCGCGCCAG 1297541
Oy	181 ThrGln 182
Db	1297542 CCGCAG 1297547
RESULT 9	
ABL05027	standard; cDNA; 3519 BP.
XX	ABL05027;
XX	ABL05027;
DT	26-MAR-2002 (first entry)
XX	
XX	Drosophila melanogaster expressed polynucleotide SEQ ID NO 9563.
DE	
XX	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical; gene; 89.
XX	

```

OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB60924.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 9563; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3519 BP; 1137 A; 643 C; 711 G; 1028 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0.00101 Length: 3519
Score: 116.00 Matches: 50
Percent Similarity: 43.07% Conservative: 37
Best Local Similarity: 24.75% Mismatches: 83
Query Match: 11.12% Indels: 32
DB: 23 Gaps: 7
US-10-069-544-2 (1-204) x ABL05027 (1-3519)
QY 18 Ser1a1CysGlnSerThrProIleProProlysAsnAsnProGlnLeuAlaGlnIleArg 37
DB 636 TCTCCCGCTTACGCTTATATACAGCGCTTACATATATATCTGATCTGTATGTCGCCG 695
QY 38 ThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeu 57
DB 696 AGCATCTGGGAAATCTTCTTACGCTCTCGTGGTTGGAAGCAAGCGCTTGTAC 755
QY 58 AspAlaIleLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrIleAlaLysVal 77
DB 756 CTAAAGCAATGAAACATGCTCGCTTGGCACTTCAGAGACTTAACCTGAGATGTGTG 815
QY 78 TyrGlnIleAspGlnAsp-----AlaThrHisGlnThrLysAlaGlnArgLeuPhe 94
DB 816 TTTAACGCTCAGGAGAAATTTGGTTGCTATACATCAC-----TTT 857
QY 95 GlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyr 114
DB 858 GAAAGAGCTGTGACTTGTATCCCAATTTTGGATCAATCAATTAATTGGGAAACGTT 917
QY 115 LeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAla 134
DB 918 CTTAAGAGAGCAACAATATTGATAGAGCTGTGCCCCCTAT-----TTACGTGCA 968
QY 135 IleGlyTyrGln---GlyArgValValAlaIleGlnAsnMetAlaTyrIleTyrTrHis 153

```

```

DB 969 TTAATCTATCTCCGAATAAAGCTGTAGTTCATGCGCAATTTGGCAATGCTTACTATGAA 1028
QY 154 Gln-----TyrGlnAlaAla-----LysSerPro 161
DB 1029 CAAGTCTTATTTAGTATGATCATACATATAGAGGGGCACTAGACATGCAACCAAT 1068
QY 162 ThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGlyThr 181
DB 1089 TTCGCCATGCTGACTGCAATCTTGGCAACGCTCTCMAA-----1127
QY 182 GlnHisAspGlnIleLysLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeuSer 201
DB 1128 GAAAGAGGCGAGCTTAAAGAGCTGAGAGCTGCTATTAACCTGACTAAGACTATGTTCA 1187
QY 202 AspTyr 203
DB 1188 AATCAT 1193
RESULT 10
AB121181
ID AB121181 standard; DNA; 3549 BP.
XX
XX AC AB121181;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 15016.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX PR 23-MAR-2000; 2000US-191637P.
XX
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX DR WPI; 2001-656860/75.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 15016; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3549 BP; 1155 A; 647 C; 715 G; 1032 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0.00103 Length: 3549
Score: 116.00 Matches: 50
Percent Similarity: 43.07% Conservative: 37

```

Best Local Similarity: 24.75% Mismatches: 83
 Query Match: 11.12% Indels: 32
 DB: 23 Gaps: 7

US-10-069-544-2 (1-204) x ABL21183 (1-3549)

```

QY 18 SerAlaCyseGlnSerThrProIleProProlYAsnAsnProGlnLeuAlaGlnIleArg 37
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 645 TCGCCGCTCAGCTATATATACAGCGTTACAAATATATCTGATCTGATTTGTCGCCG 704
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 38 ThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeu 57
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 705 AGCGATCTGGGAATCTTCTTAAGCTCTCGCTGCTTGAAGAAAGCAAGCGCTGTGTAC 764
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 58 AspAlaIleLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysVal 77
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 765 CTAAAGCAATTGAAACATGCTCGGCTTTGCACTTCAGAGTAACCTAGATGTGTG 824
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 78 TyrGlnAlaSerGlnAsp-----AlaThrHisGlnThrLysAlaGlnArgLeuPhe 94
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 825 TTTAAGCTCAGGAGAAATTTGGTTGGCTATACATCAC-----TTT 866
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 95 GlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyr 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 867 GAAAAAGCTGTGACTCTTGATCCCAATTTTGGATGCATATAAATTTGGGAAAGCTT 926
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 LeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAla 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 927 CTTAAGAGCAACCAATATTTGATAGAGCTGTGCGCCCTAT-----TTAGGTCA 977
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 135 IleGlyTyrGln---GlyArgValValAlaIleGlnAsnMetAlaTyrIleTyrHis 153
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 978 TTAATCTATCTCCGATATATGCTGTAGTTCAGGCAATTGGATCGTTACTATGAA 1037
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 154 Gln-----TyrGlnAlaAla-----LysSerPro 161
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1038 CAAGCTCTTATTTAGCTATGATCATATATAGAGGCAATAGAACTGCAACCAAT 1097
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 162 ThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGlyThr 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1098 TTCCCGATCGTACGCAATCTTGCAACGCTCTCAA----- 1136
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 182 GlnHisAspGlnIleLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeuSer 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1137 GAAAGGGGCAAGCTTAAGAAGCTGAGACTGCTATTAACACTGCACTAAGCTATGTTC 1196
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 202 AspTyr 203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1197 AATCAT 1202
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 11

```

ID ABL21183 standard; DNA; 3604 BP.
XX ABL21183;
XX
AC ABL21183;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 15022.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; de.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
PR

```

XX (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PMD, Myers EM;
 XX WPI; 2001-656860/75.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Claim 1; SEQ ID NO 15022; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
 CC sequences (AB85737-AB87072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

CC Sequence 3604 BP; 1172 A; 654 C; 720 G; 1058 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Score:
0.00105	3604	116.00
Percent Similarity:	43.07%	Conservative: 37
Best Local Similarity:	24.75%	Mismatches: 83
Query Match:	11.12%	Indels: 32
DB:	23	Gaps: 7

US-10-069-544-2 (1-204) x ABL21183 (1-3604)

```

QY 18 SerAlaCyseGlnSerThrProIleProProlYAsnAsnProGlnLeuAlaGlnIleArg 37
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 700 TCGCCGCTCAGCTATATATACAGCGTTACAAATATATCTGATCTGATTTGTCGCCG 759
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 38 ThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeu 57
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 760 AGCGATCTGGGAATCTTCTTAAGCTCTCGCTGCTTGAAGAAAGCAAGCGCTGTGTAC 819
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 58 AspAlaIleLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysVal 77
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 820 CTAAAGCAATTAAGAAATGCTCGGCTTGGCAGTTGAGAGTAACCTAGAGATGTGTG 879
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 78 TyrGlnAlaSerGlnAsp-----AlaThrHisGlnThrLysAlaGlnArgLeuPhe 94
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 880 TTTAAGCTCAGGAGAAATTTGGTGGCTATACATCAC-----TTT 921
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 95 GlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyr 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 922 GAAAGGCTGTGACTGTATCCCAATTTTGGATGCATATAAATTTGGGAAAGCTT 981
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 LeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAla 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 982 CTTAAGAGGCAAGAAATTTGATAGAGCTGTGCGCCCTAT-----TTACGTCA 1032
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 135 IleGlyTyrGln---GlyArgValValAlaIleGlnAsnMetAlaTyrIleTyrHis 153
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1033 TTAATCTATCTCCGATATATGCTGTAGTTCAGGCAATTGGCAAGCGCTTACTATGAA 1092
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 154 Gln-----TyrGlnAlaAla-----LysSerPro 161
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1093 CAAGCTCTTATTTAGCTATGATCATATATAGAGGCAATAGAACTGCAACCAAT 1152
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 162 ThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGlyThr 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1153 TTCCCGATCGTACGCAATCTTGCAACGCTCTCAA----- 1191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

QY 182 GluHisAspGluIleLysLysSerTyrAspLysLeuSerAspTyrLysLeuSer 201
Db 1192 GAAAGGGGCGAGTTAAAGAGCTGAGACTGCTATAACACTGCACTAAGACTATGTTCA 1251
QY 202 AspTyr 203
Db 1252 AATCAT 1257
RESULT 12
AAZ54212
ID AAZ54212 standard; DNA; 762 BP.
XX
AC AAZ54212;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 708 partial DNA sequence SEQ ID NO:2373.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy; ds.
XX
OS Neisseria gonorrhoeae.
XX
PN WO957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Peteren J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tetrelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; AAY75450.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 1140; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 762 BP; 213 A; 219 C; 173 G; 157 T; 0 other;
Alignment Scores: 0.000333 Length: 762
Pred. No.: 112.50 Matches: 42
Score: 112.50

Percent Similarity: 46.70% Conservative: 43
Best Local Similarity: 23.08% Mismatches: 78
Query Match: 10.79% Indels: 19
DB: 21 Gaps: 5
US-10-069-544-2 (1-204) x AAZ54212 (1-762)
QY 4 ArgValLysTrpPrometValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThr 23
Db 22 CGATCTCTTTATTACTGTCTTCTGCTGGC-----GGTCCAGCACTTCC 69
QY 24 -----ProIleProProlLysAsnAspProGlnLeuAlaGlnIleArgThrGlnIleAla 41
Db 70 TACCGCCCTCGGGGAGAAAAGCAATCAGGTTCCAAATATCAAAACCGATTGGCC 129
QY 42 IleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeu 61
Db 130 ATGGATATATGCGCGGTACGACTACCGTCAGGCAACGCAAGTATTTGAAGATGCTTG 189
QY 62 SerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSer 81
Db 190 AAATCGAACCCCTAAAAACGAACCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249
QY 82 GluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsn 101
Db 250 AAA-----GTTAACGACAAAGCGCGCAAGAAAGTTCCGCGCAAGCCTTCATCAAA 300
QY 102 ProlLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuVal---GlnMetGlyAsp 120
Db 301 CCCGACACGTCGCGAAATCAACAACAACTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 121 LeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGlnGlyArg 140
Db 361 CTTGCCGAATCTATGCGATATTTGCAAAAGCCCTGCGCAGCCCACTTCCGAGCCCT 420
QY 141 ValValAlaIleGlnLysMetAlaTyrIleTyrThrHisGlnThrGlnAlaAlaLysSer 160
Db 421 TATATTGCCAAGCTGAAT-----AAAGTATATGAGC 453
QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGly 180
Db 454 GCAAAACAGGGCGCAATTCGATTGCGGGAAGCCTATTGAAAGTTCCCTGCGCCAG 513
QY 181 ThrGln 182
Db 514 CCGCAG 519
RESULT 13
ABN91309
ID ABN91309 standard; DNA; 693 BP.
XX
AC ABN91309;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:772.
XX
KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KM antibacterial; gene therapy; gene; ds.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.

DB 794 ACAAGTGTAGAGCTGAGCGTGAAGAACTCTTCTTGAAAGCTATTGAGCTGATCCACC 853
QY 104 ASPMERCINSEPTYRMEASPTYRGILYPHEITYRLEUVALGINMETGLYASPLEUSERGILY 123
DB 854 AAGGAAACGTGTATCATGCATTATGTCAGTTCCTCTCGAAA----- 895
QY 124 ALALEULLEYRPHENAPLYSPROSERARGALALEULLEYRGLUGLYARGVALVALA 143
DB 896 -----GAAAGCTGCTTCATGAA 913
QY 144 ILGLUASMETALATYRILEYRITYRHISGLINTYRGILUALALYS-----SERPRO 161
DB 914 GCAGCTGAGATGGA-----AAAAAGAGCTGAACTGAACTGACAGACA 955
QY 162 THIRLSAPSPAPTYRANASNALALYSERIALALEUGLYARGALALEULLESERGLYTHR 181
DB 956 GAGTTTGATGTGCTTCATGCTGACCCACATGCTCAGACGCTGACCTC----- 1006
QY 182 GLNHISAPGLULEYLSYSEPTYRASP 191
DB 1007 ---AATGACGACGCTGAGAAATATTATGAT 1033
RESULT 15
AAH54213/c
ID AAH54213 standard; DNA; 4241 BP.
XX AAH54213;
AC
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3577.
XX
KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAXO) GLAXO GROUP LTD.
XX
PI Kimerly WJ;
XX
DR WPI; 2001-316495/33.
PT Nucleic acid encoding polypeptides from Staphylococcus epidermidis,
PS useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1164-1166; 2188pp; English.
XX
CC AAH52304 to AAH3970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH3971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH5091 to
CC AAH5098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,

CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 4241 BP; 1541 A; 676 C; 578 G; 1446 T; 0 other;
Alignment Scores:
Pred. No.: 0.0193 Length: 4241
Score: 106.50 Matches: 36
Percent Similarity: 46.56% Conservative: 25
Best Local Similarity: 27.48% Mismatches: 53
Query Match: 10.21% Indels: 17
DB: Gaps: 5
US-10-069-544-2 (1-204) x AAH54213 (1-4241)
QY 75 ALALYSVALTYRGLNALASERGLUASPALATHRHISGLINTYRGLNALARGLEUPHE 94
DB 761 AGTAAGATATACCAAAATATTAAGATGTAATTAAGATACCGCTTGAAGACCTGTTT 702
QY 95 GLULYSALAILGLULEUASNPOLYSASPMETGLINSEPTYRMEASPTYRGILYPHEITYR 114
DB 701 GAAAT--ATGAAAGAAAATCCGGCCATGTGTAATAATTAATTAACCGAGTATTGTA 645
QY 115 LEUVALGIMERCGLYASPLEUSERGLYALALEULLEYRPHENAPLYSPROSERARGALA 134
DB 644 TTRATGATGTTGGAGAAATTTGAAAAAGCAGAACGTTTTCAAAAGGCATTAACTATT 585
QY 135 ILGLYTYRGLUGLYARGVALVALALAILGLUASMETALATYRILEYRITYRHISGLIN 154
DB 584 GAACCTGAAAATGCTGCTGT-----TATTATTAACCTTGCAAAATTTATTAATATGAA 531
QY 155 TYRGLUALALEULYSERPROTHIRLSASPSPTYRANASNALALYSERIALALEUGLY 174
DB 530 -----GAAAGCTGCTTCATGAA 194
QY 175 ARGALALEULLESERGLYTHRGILNHISAPGLULEYLSYSEPTYRASPILYSEUL 194
DB 497 ACTGCACCTA-----CAATATGAAGTAGCAAAAAGATGTATATATATGATT 450
QY 195 SER--ASPTYRILYSLEULEUSERSPITYRILYS 204
DB 449 GGTATGTCGTTTATCATCACTAGTGTCTTTAAA 417
RESULT 16
ABO67195
ID ABO67195 standard; DNA; 495269 BP.
XX
AC ABO67195;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua contig DNA sequence #8.
XX
KM Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.

XX New genomic sequences from *Listeria* species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators -
 XX
 PS Claim 5; SEQ ID 8; 180bp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes
 CC and primers for identification and/or detection of *Listeria* (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
 CC treating infections by *Listeria*, and are useful as immunogens in
 CC anti-*Listeria* vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 495269 BP; 159529 A; 88849 C; 93135 G; 153754 T; 2 other;
 XX
 Alignment Scores:
 Pred. No.: 13.4 Length: 495269
 Score: 106.50 Matches: 43
 Percent Similarity: 37.79% Conservative: 22
 Best Local Similarity: 25.00% Mismatches: 66
 Query Match: 10.21% Indels: 41
 DB: 24 Gaps: 4
 US-10-069-544-2 (1-204) x AB067195 (1-495269)
 QY 44 LeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaLysSerAla 63
 DB 211562 TTAATTTTCATGTTAGAAATGTTTGAACGGCGAATTTATTTTCAAAAGCGCTTGAAC 211621
 QY 64 AspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyr 78
 DB 211622 GATGACACGGTCCGGCGCTTATTTAGCTTAGTAATTTATTTAGCAATTAGAACGT 211681
 QY 79 -----GlnAlaSerGluAspAlaThrHisGln----- 87
 DB 211682 TATCAAGAAAGCTCGCGATAGCTTTCAAAATGCAACAAAGAAATGGAATGGCGAT 211741
 QY 88 -----ThrLysAlaGlnArg 92
 DB 211742 TTAATTTTCATGTTAGAAATGAGTTTGTTCAAATGAGAGATTAAACGCTAGTATGCT 211801
 QY 93 LeuPheGluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112
 DB 211802 TATTTACTTGAAGTGTGCAATTAATCCAAAGATAGTGAACGCTTATTTCAATATGCT 211861
 QY 113 PheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSer 132
 DB 211862 ATCGTTCTTCTCGCAGCGGA-----TTTAAACGAAGAT----- 211894
 QY 133 ArgAlaIleLeuTyrGluGlyArgValAlaIleGluLeuMetAlaTyrIleTyrTyr 152
 DB 211895 ---GCGATTAATATGCTTGAAGAGTACTTCTTATTAACAAAGATCCAGATGCACCT 211951
 QY 153 HisGlnTyrGluLysAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSerAla 172
 DB 211952 TATTAATATCGAGCTGTTATTCGGCATGCGCAAGCGACATTTGACTTGCAGAAAATTTAC 212011
 QY 173 LeuGluArgAlaLeuIleSerGlyThrGlnHisAsp 184
 DB 212012 TTGAAACGCGCGCTGCACTGCAATGCAATCCCATGAA 212047
 RESULT 17
 AB069245/c
 ID AB069245 standard; DNA; 3011208 BP.
 XX
 AC AB069245;

XX
 DT 29-AUG-2002 (first entry)
 XX
 DE *Listeria innocua* DNA sequence #684.
 XX
 KW Antibacterial; *Listeria*; food contamination; mutational analysis;
 KW infection; ds.
 XX
 OS *Listeria innocua*.
 XX
 PN MO200228691-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 04-OCT-2001; 2001WO-FR03061.
 XX
 PR 04-OCT-2000; 2000PR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kuner F, Glaeser P;
 XX
 DR WPI; 2002-332479/37.
 XX
 PT New genomic sequences from *Listeria* species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators -
 XX
 PS Claim 5; SEQ ID 2058; 180bp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes
 CC and primers for identification and/or detection of *Listeria* (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
 CC treating infections by *Listeria*, and are useful as immunogens in
 CC anti-*Listeria* vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 160 Length: 3011208
 Score: 106.50 Matches: 43
 Percent Similarity: 37.79% Conservative: 22
 Best Local Similarity: 25.00% Mismatches: 66
 Query Match: 10.21% Indels: 41
 DB: 24 Gaps: 4
 US-10-069-544-2 (1-204) x AB069245 (1-3011208)
 QY 44 LeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaLysSerAla 63
 DB 1547393 TTAATTTTCATGTTAGAAATGTTTGAACGGCGAATTTATTTTCAAAAGCGCTTGAAC 1547334
 QY 64 AspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyr 78
 DB 1547333 GATGACACGGTCCGGCGCTTATTTAGCTTAGTAATTTATTTAGCAATTAGAACGT 1547274
 QY 79 -----GlnAlaSerGluAspAlaThrHisGln----- 87
 DB 1547273 TATCAAGAAAGCTCGCGATAGCTTTCAAAATGCAACAAAGAAATGGAATGGCGAT 1547214
 QY 88 -----ThrLysAlaGlnArg 92
 DB 1547213 TTAATTTTCATGTTAGAAATGAGTTTGTTCAAATGAGAGATTAAACGCTAGTATGCT 1547154
 QY 93 LeuPheGluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112

XX differential gene expression; genomic sequenced tag; GST;
 KM altered culture condition; environmental stress;
 KM physiological provocation; ds.
 OS *Bacillus licheniformis*.
 PN WO200229113-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 05-OCT-2001; 2001WO-US31437.
 XX
 PR 06-OCT-2000; 2000US-0680598.
 PR 27-MAR-2001; 2001US-279526P.
 XX
 PA (NOVO) NOVOZYMES BIOTECH INC.
 PI (NOVO) NOVOZYMES AS.
 XX Berka R, Clausen IG;
 PI
 XX WPI; 2002-416684/44.
 DR
 XX
 PT Monitoring differential expression of several genes in first *Bacillus*
 PT cell relative to expression of same genes in one or more second
 PT *Bacillus* cells, by using substrate containing *Bacillus* genomic
 PT sequenced tag array -
 XX
 PS Claim 4, SEQ ID NO 1935; 200pp; English.
 XX
 CC The invention describes a method of monitoring differential expression of
 CC genes in a first *Bacillus* cell relative to expression of the genes in
 CC other *Bacillus* cells, comprising hybridising labelled nucleic acid probes
 CC isolated from *Bacillus* cells to a substrate containing array of *Bacillus*
 CC genomic sequenced tags (GST), examining the array, and determining
 CC relative gene expression by an observed hybridisation reporter signal of
 CC a spot in the array. The method is useful for measuring the expression of
 CC genes in a first *Bacillus* cell relative to expression of the same genes
 CC in one or more second *Bacillus* cells. The method is useful for monitoring
 CC global expression of several genes from a *Bacillus* cell, discovering new
 CC genes, identifying possible functions of unknown open reading frames and
 CC monitoring gene copy number variation and stability. Monitoring changes
 CC in expression of genes may be used to provide a representation of the way
 CC in which *Bacillus* cells adapt to changes in culture conditions,
 CC environmental stress or other physiological provocation. Extensive
 CC follow-up characterisation is unnecessary, when one spot on an array
 CC equals one gene or one open reading frame, since sequence information is
 CC available. This sequence represents a genomic sequence tag (GST) used in
 CC the method of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1971 BP; 604 A; 470 C; 505 G; 390 T; 2 other;
 XX
 Alignment Scores:
 Pred. No.: 0.00774 Length: 1971
 Score: 106.00 Matches: 47
 Percent Similarity: 41.71% Conservative: 31
 Best Local Similarity: 25.13% Mismatches: 73
 Query Match: 10.16% Indels: 36
 DB: 24 Gaps: 8
 US-10-069-544-2 (1-204) x ABK74644 (1-1971)
 QY 18 SerAlaCysGlnSerThrProIleProIleAsnAsnProGlnLeuAlaGlnIleArg 37
 DB 219 ACAGCTCGCGCGTCCGTCGCGAGCGCGTAAAAA-----CAGCGATCAAGTCGCG 272
 QY 38 ThrGlnIleAlaIleSerLeuAspMetCysGlyLeuLeuAspGlnAlaIleGlnIleu 57
 DB 273 AAAGACATCGCATCGAACCAATTAAACAGCAGCAAAAGTATCGGCTGTCTTGA AAAA--- 329

QY 58 AspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaIleVal 77
 DB 330 ---GGGGGATGCGACAGCTTCATCAGCGCGCCGAAACGTTACATCTAAATATG 386
 QY 78 TyrGlnAlaSerGlnAspAlaThrHseGlnThrTyrAlaGlnArgLeuPheGlnIleVal 97
 DB 387 AAAGAAAGCAAAAACCGGATCTTCAATCTAAAGCCCAATCATCCGCGCTCCAGCA 446
 QY 98 IleGlnLeuAsnProIleAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGln 117
 DB 447 TTAGAGCCCTCGTAAAGCAAAACGGATTAAAGT-----ATGATCCTTGGAA 500
 QY 118 MetGlyAspLeuSerGlyValaLeuIleTyrPheAspTyrProSerArgAlaIleGly--- 136
 DB 501 ATG-----AGCAAGCCCTCTCCGATTC 527
 QY 137 -----TyrGlnIleArgValValAlaIleGlnAsnMetAlaTyrIle 150
 DB 528 AAAGCGCGCAAGCAATATCATCGCGCGCCTCCCAAGATGCCAAGATGCTCGAAGTG 587
 QY 151 TyrTyr-----HisGlnTyrGlnAlaAlaIleAspSerProThrTyrAspAspTyr 166
 DB 588 GAATTCGTGCAACAGTTCAGCAATATGAGCACTGTCAAGACACCCCAATATCCATAT 647
 QY 167 -----AsnAsnAlaIleSerAlaLeuGlnArgAlaLeuIleSerGlyThr 181
 DB 648 CAATGTCGCTCAAAAATATACGCAAAAC-----CTGTCTCGCAATCTGACATA 698
 QY 182 GlnHisAspGlnIleIleValLeu 188
 DB 699 CAATTGAACAGCTTCAGAG 719
 RESULT 20
 AA254214
 ID AA254214 standard; DNA; 762 BP.
 AC AA254214;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE *Neisseria meningitidis* ORF 708 partial DNA sequence SEQ ID NO:2377.
 XX
 KM *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy; ds.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN WO957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121526.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzo M, Rappelli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 P-PSDB; AA75452.


```

Db      750093  AAGATGAGAAATTAATAAGCACTTTTGGATTGGAAAGGCTATTATTAAATGAGT 750034
Qy      83      ACPALAThrrIglInthrIysAlaIglInArgLeuPheGluLysAlaIleGluLeuAanPro 102
Db      750033  GATTAATAAAATTCATTAATAA-----TACTTGGAAAATACTTTAGATTAATAATCCA 749983
Qy      103     LysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSer 122
Db      749982  AATGATGTAGAGCA---TTGGAAATAT-----TTG 749956
Qy      123     GlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArgValVal 142
Db      749955  GGAGGCTTATATTATGACAGAAAGATTGTGAAAAGCCATTAACCTTTAAAGGCTTAA 749896
Qy      143     AlaIleGlu-----AsnMetAlaTyrIleTyrIleGln 154
Db      749895  GAATTAAACCCAGATGATATTGATTAAATTTTAAAGTTGCTTACATTACTTCAAAATTG 749836
Qy      155     TyrGluAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlu 174
Db      749835  AAGAA-----TATTAACATGCACCTAAATTTTGGAG 749803
Qy      175     ArgAlaLeuIleSerGlyThrGlnHisAspGluIleLysLysSerTyrAspLysLeuLeu 194
Db      749802  AAGCTTTAAATTAATTAATCCAAATGCTTTGAATTAGAACACATTAATCTATGCGA 749743
Qy      195     SerAspTyrLysLeuLeuSerAsp 202
Db      749742  AGAATATACATTTATTAGAGAA 749719

```

RESULT 22
ABAO3041/c
ID ABAO3041 standard; DNA; 2944528 BP.

```

XX      ABAO3041;
XX      05-FEB-2002 (first entry)
XX      Listeria monocytogene EGD-e genome sequence.
XX      Antibiobacterial; gene therapy; vaccine; bioenhancers; biodegradation;
XX      vitamin B12; bacterial infection; disease; de.
XX      Listeria monocytogenes.
XX      MO200177335-A2.
XX      18-OCT-2001.
XX      11-APR-2001, 2001WO-FR01118.
XX      11-APR-2000, 2000FR-0004629.
XX      (INSP) INST PASTEUR.
XX      Buchrieser C, Frangoul L, Couve E, Ruenick C, Feibi H, Dehoux P,
XX      Dussanrgot O, Chetoui F, Nedjari H, Glaeser P, Kunet P, Cosserat P,
XX      Daniels J, Goebel W, Kreft J, Kuhn M, Ng B, Vazquez-Boland JA,
XX      Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
XX      Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L,
XX      Peres-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
XX      Medunio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J,
XX      Rose W, Voss H,
XX      WPI, 2002-010914/01.
XX      Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX      and prevention of Listeria and related bacterial infections, and
XX      related polypeptides
XX      Claim 1; SEQ ID No 1, 192pp; French.

```

CC The present sequence is the genome sequence of Listeria monocytogenes
CC EGD-e. This sequence and fragments of this sequence are useful for
CC selecting probes and primers for detecting genes in L. monocytogenes and
CC related organisms, and to study genetic polymorphisms and other genomes.
CC Protein (AB847297-AB850149) expressed from the present sequence are
CC useful for raising specific antibodies, identification of L.
CC monocytogenes and related organisms, and for biosynthesis and
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC proteins encoded by it are also useful for selecting compounds that
CC regulate gene expression and cell replication and modulate L.
CC monocytogenes-related diseases. In addition, this sequence and proteins
CC encoded by it are useful in pharmaceutical and vaccine compositions for
CC the treatment or prevention of infections by L. monocytogenes and related
CC organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Alignment Scores:
Pred. No.: 638 Length: 2944528
Score: 101.50 Matches: 41
Percent Similarity: 37.79% Conservative: 24
Best Local Similarity: 23.84% Mismatches: 66
Query Match: 9.73% Indels: 41
DB: 24 Gaps: 4

US-10-069-544-2 (1-204) x ABAO3041 (1-2944528)

```

Qy      44     LeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaLysSerAla 63
Db      1542942  TTACTTCAATGATGATGATTTTGAAGTGGGAAATATTATTTCAAAAGCTCTTGAGCTA 1542883
Qy      64     AspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyr----- 78
Db      1542882  GACGACACAGTTCACGCTGCTTACTTATAGCTTACGAAATTTATATGATTAAGTACGCT 1542823
Qy      79     -----GlnAlaSerGlnAspAlaThrHisGln----- 87
Db      1542822  TATCAAGAAAGCCGACGATAGCTTCCAAATATGCCACAAACAGAAATGGAATGCTGAT 1542763
Qy      88     -----ThrLysAlaGlnArg 92
Db      1542762  TTGTTTTCATGTTAGATGATGTTTGTTCAAATGAGAGACTCAACTTCGATGCCA 1542703
Qy      93     LeuPheGluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112
Db      1542702  TATTTACTTAAAGTGTGCAATTAATCCAGAAAGCGGCAAGCGCTTATTTCAATATAGGG 1542643
Qy      113     PheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSer 132
Db      1542642  ATTGTGCTTGTCTGCGAGTGT-----TTTATGAAAT----- 1542610
Qy      133     ArgAlaIleGlyTyrGluGlyArgValAlaAlaIleGluAsnMetAlaTyrIleTyr 152
Db      1542609  ---GGATTAATATGTTAGAACGCTTTTACTGTTTACCGCCGAGATCCAGATGACACT 1542553
Qy      153     HisGlnTyrGluAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSerAla 172
Db      1542552  TACCAATATTGGTGCGACTTATGCGATGCGCAAGAGATATGCTTCTGCTTAAAACTAT 1542493
Qy      173     LeuGluArgAlaLeuIleSerGlyThrGlnHisAsp 184
Db      1542492  TTTGAACGACGCAATTGCTACAGCGCATCACAATGA 1542457

```

RESULT 23
AAV05171
ID AAV05171 standard; cDNA; 3476 BP.
XX AAV05171;
XX 19-MAY-1998 (first entry)

Query Match: 9.49% Indels: 34
 DB: 20 Gaps: 6
 US-10-069-544-2 (1-204) x AAX20551 (1-6993)

QY 41 AAlaIleSerLeuLeuAspMetGlyLyLeuAspGlnAlaLeuGlnLeuAspAlaAla 60
 DB 5799 GCGTTCAGTTAGTTGTCACAGTGAAGTTCAGGCGCCGTTGTTTACAGGAATT 5740
 QY 61 LeuSer-----AlaAspAGGlnPheAlaProAlaTyrTrpGhrLeuAla 75
 DB 5739 CTTGGCGCTATCCGAATGACCGATGCTCGCTTTGTTGGTTCAGAGATTGAGTGTA 5680.
 QY 76 LysValaIYrGlnAlaSerGlnAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGln 95
 DB 5679 AAGGACGCGCTGAGCT-----GCGGCTCTGTTATACCA 5644
 QY 96 LysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeu 115
 DB 5643 GCAGCGCTGAGCGTCAGCGCAGAGATCCCAAGCGCTTGTGCGCTTCGCTTATCTCG 5584
 QY 116 ValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIle 135
 DB 5583 TACGAGCGCTGTCATTATCCGCGCTGCTCAGCTACGTTGAA-----CGTGGCTTG 5533
 QY 136 GlyTyrGlnGlyArgValAlaAlaIleGlnAsnMetAlaTyrIleTyrTrpHisGlnTyr 155
 DB 5532 CAATATCAGCGTGAATATGACACAGTG-----CATTTT 5497
 QY 156 GluAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArg 175
 DB 5496 GCGCGCTATCTTCAACGCTGCGCACACTATGAAGAAGCAGACGCTTATTTGAGCGG 5437
 QY 176 AlaLeuIleSerGlyThrGlnHisAspGlnIleLysLysSerTyrAspLys----- 192
 DB 5436 GGGCTC-----CACATTAACTCGCTTATCCAGTCCCGCTGCG 5398
 QY 193 LeuLeuSer 195
 DB 5397 CTTCTTTCA 5389

RESULT 25
 AAV69301
 ID AAV69301 standard; DNA; 3083 BP.
 AC AAV69301;
 XX
 DT 01-FEB-1999 (first entry)
 XX
 DE Human OGT DNA.
 XX
 KM OGT, O-linked GlcNAc transferase; uridine; transferase; human; tumour;
 KM dihydroxy-N-acetylglucosamine; polypeptide beta-N-acetylglucosaminyl;
 KM predilection; type II diabetes; hyperglycaemia; Alzheimer's disease;
 KM metastasis; diagnosis; inhibitor; treatment; diabetes mellitus; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 265..3027
 FT /tag= a
 FT /product= "OGT"
 FT /note= "O-linked GlcNAc transferase"
 XX
 PN W09044123-A2.
 XX
 XX 08-OCT-1998.
 XX PD
 XX 27-MAR-1998; 98WO-US06101.
 XX PF
 XX 31-MAR-1997; 97US-0042270.
 XX PR
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA

XX
 PI Hanover JA, Lubas W;
 XX
 DR MPI: 1998-557118/47.
 XX
 DR P-PSDB; AAM82500.
 XX
 PT Protein exhibiting O-linked GlcNAc transferase activity, OGT -
 PT useful, e.g. to assess predisposition to type II diabetes or
 PT Alzheimer's or metastatic potential of tumours, and to identify
 PT inhibitors
 XX
 PS Claim 4, Page 27-31, 56pp; English.
 XX
 CC This sequence encodes a novel human O-linked GlcNAc transferase, OGT
 CC protein (also known as uridine dihydroxy-N-acetylglucosamine:
 CC polypeptide beta -N-acetylglucosaminyl transferase). This protein is
 CC useful to assess predisposition toward type II diabetes in patients
 CC suspected of having hyperglycaemia that could evolve into this disease,
 CC by assaying OGT activity in red blood cells. It can also be used to
 CC assess predisposition toward Alzheimer's disease, to assess the
 CC metastatic potential of tumours and to diagnose a tumour with metastatic
 CC potential. OGT can also be used to identify OGT inhibitors, especially in
 CC high-throughput assays, useful, e.g. in the treatment of diabetes
 CC mellitus, tumour-derived diseases and Alzheimer's disease.
 XX

Sequence 3083 BP; 859 A; 690 C; 699 G; 835 T; 0 other;

Alignment Scores:
 Pred. No.: 0.119 Length: 3083
 Score: 98.50 Matches: 52
 Percent Similarity: 40.31% Conservative: 27
 Best Local Similarity: 26.53% Mismatches: 82
 Query Match: 9.44% Indels: 36
 DB: 19 Gaps: 8

US-10-069-544-2 (1-204) x AAV69301 (1-3083)

QY 17 LeuSerAlaCysGlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIle 36
 DB 842 CTCCTCCGCTGTGT-CCACCCATGACAGACTCTGAAATAC-----CTAGCCATATTC 894
 QY 37 ArgThrGlnIleAlaIleSerLeuAspMetGlyLyLeuAspGlnAlaLysGlnGln 56
 DB 895 AAACGA-----GACACGGAACATGAGAGACGACGCTTG 933
 QY 57 LeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLys 76
 DB 934 TATCGTAAAGCATTTGAAGTCTTCCACAGATTTGCTGCTGCCCATTCAAATTTAGCAAGT 993
 QY 77 ValTyrGlnAlaSerGlnAspAlaThrHisGlnThrLysAlaGlnArgLeu----- 93
 DB 994 GTATCTGCG-----CACAGGGAACATGACGAAGCTTGATGCT 1035
 QY 94 PheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPhe 113
 DB 1036 TATAGAGGCGTATTCGATCAGTCTTCTGCTGATGCTTCTTCAATATGGAAC 1095
 QY 114 TyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArg 133
 DB 1096 ACTTAAAGAGATCCAGATGTTTCAGGAGGCTTGACGTGTAT-----ACGCGT 1146
 QY 134 AlaIleGlyTyrGlnGlyArgValAla-----AlaIleGlnAsnMetAlaTyrIleTyr--- 151
 DB 1147 GCCATCCAAATTTATCTCGCATTTTCCAGATGACATGACATCTGCTTCCATTCATAG 1206
 QY 152 -----TyrHisGlnTyrGlnAlaAlaLysSer 160
 DB 1207 GATTCAGGAATATTCAGAGCCATAGCTTCTTACCGACGCGCTGAAACTTAAGCT 1266
 QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGly 180
 DB 1267 GATTTCTCGTATGTTGTAATCTTGATGCTCATGCTGCGAG-----ATTGCTGTGATTCG 1323

QY 181 ThrGlnHisAspGluIleIysSerTyrAspIysLeuSerasp 196
 DB 1324 ACAGACTATGATGAGCAATGAAGATTGTGTGATTTGTGCTGAC 1371
 RESULT 26
 ID ABR84430 standard, cDNA; 3084 BP.
 XX ABR84430;
 AC ABR84430;
 DT 14-AUG-2002 (first entry)
 XX
 XX Human cDNA differentially expressed in granulocytic cells #1001.
 DE Human; ss; granulocytic cell; DNA chip; bacterial infection;
 XX viral infection; parasitic infection; protozoal infection;
 KM fungal infection; sterile inflammatory disease; psoriasis;
 KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KM cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KM adult respiratory distress syndrome; inflammatory bowel disease;
 KM Crohn's disease; ulcerative colitis; periodontal disease;
 KM granulocyte activation; chronic inflammation; allergy.
 XX Homo sapiens.
 OS WO200228999-A2.
 PN 11-APR-2002.
 PD 03-OCT-2001; 2001WO-US90821.
 PF 03-OCT-2000; 2000US-237189P.
 PR 03-OCT-2000; 2000US-237189P.
 XX
 XX (GENE-) GENE LOGIC INC.
 PA Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 PI WPI; 2002-435328/46.
 DR
 XX
 PT Detecting granulocyte activation by detecting differential expression
 of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 PS Claim 1, SEQ ID No 1001; 114BP; English.
 XX
 XX The invention relates to detecting (M1) granulocyte (GC) activation
 (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,

CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcc_sequences.
 XX
 SQ Sequence 3084 BP; 859 A; 690 C; 699 G; 836 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.119 Length: 3084
 Score: 96.50 Matches: 52
 Percent Similarity: 40.31% Conservative: 27
 Best Local Similarity: 26.53% Mismatches: 82
 Query Match: 9.44% Indels: 36
 DB: 24 Gaps: 8
 US-10-069-544-2 (1-204) x ABR84430 (1-3084)
 QY 17 LeuSerAlaCysGlnSerThrProIleProPolsAsnAsnProGlnLeuAlaGlnIle 36
 DB 843 CTCCTCGTGTGTGT-CCCAACCAGCAGACTCTCTGAATTAAC-----CTAGCCAAATATC 895
 QY 37 ArgThrGlnIleAlaIleSerLeuAspMetGlyIysLeuAspGlnAlaGlnGln 56
 DB 896 AAACGA-----GAAACGGAACAACTTGAAGAGCAGCTTGCTTG 934
 QY 57 LeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaIys 76
 DB 935 TATCGTAAGCATTTAGAACTCTCCCAAGATTGCTGCTGCCCATTTCAATTACCAAGT 994
 QY 77 ValTyrGlnAlaSerGlnAspAlaThrHisGlnThrIysAlaGlnArgLeu----- 93
 DB 995 GTACTGCAG-----CAGCAGGAAACCTGAGAGAGCTGTGATGCAT 1036
 QY 94 PheGlnIysAlaIleGlnLeuAsnProIysAspMetGlnSerTyrMetAspTyrGlyPhe 113
 DB 1037 TATTAAGAGGCTATTGAATCAGTCTTACTTGTGATGCTTACTTAAATGGAAC 1096
 QY 114 TyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspIysProSerArg 133
 DB 1097 ACTCTTAAGAGATGAGATGAGATTTTCAGGAGCTTGCACTGTAT-----ACGCGT 1147
 QY 134 AlaIleGlyTyrGlnGlyArgValVal---AlaIleGlnAspMetAlaTyrIleTyr--- 151
 DB 1148 GCCATCCAAATTAACTCTGATTTGCAGATGCACATAGCAATCTGCTTCATTCATTAAG 1207
 QY 152 -----TyrHisGlnTyrGlnAlaIleAspSer 160
 DB 1208 GATTCAAGGAATATTCACAGACCAATAGCTTTTACCGACGGCTTGAAACTTAAGCT 1267
 QY 161 ProThrIysAspAspTyrAsnAsnAlaIysSerAlaLeuGlnAlaLeuIleSerGly 180
 DB 1268 GATTTTCCGATGCTATTATGTAACCTTGCTCATGCTCAG---ATTGCTGTGATGG 1324
 QY 181 ThrGlnHisAspGluIleIysSerTyrAspIysLeuSerasp 196
 DB 1325 ACAGACTATGATGAGCAATGAAGATTGTGTGATTTGTGCTGAC 1372
 RESULT 27
 ID ABR96929 standard, DNA; 3084 BP.
 XX ABR96929;
 AC ABR96929;
 DT 13-AUG-2002 (first entry)
 XX
 XX Gene #3427 used to diagnose liver cancer.
 DE Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;

KW	metastatic liver tumour; cytoskeletal; expression profile; disease state;
KV	disease progression; drug toxicity; drug efficacy; drug metabolism.
XX	
XX	
OS	Homo sapiens.
PN	WO200229103-A2.
XX	
PD	11-APR-2002.
PF	02-OCT-2001; 2001WO-US30589.
XX	
PR	02-OCT-2000; 2000US-237054P.
XX	
PA	(GENE-) GENE LOGIC INC.
XX	
PI	Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX	
DR	WPI; 2002-426119/45.
XX	
FT	Diagnosing and detecting the progression of liver cancer,
FT	hepatocellular carcinoma or metastatic liver tumor in a patient,
PT	involves detecting the level of expression of two or more genes in a
PT	liver tissue sample -
XX	
PS	Claim 1, SEQ ID NO 3427, 298pp, English.

	CC	The invention relates to a novel method for diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma or metastatic liver
	CC	tumour in a patient, and differentiating metastatic liver cancer from
	CC	hepatocellular carcinoma in a patient, involving detecting the level of
	CC	expression of two or more genes represented in ABN93503-ABN97455 in a
	CC	tissue sample. The method of the invention has hepatotropic, and
	CC	cycostatic activity. The method is useful for diagnosing and detecting
	CC	the progression of liver cancer, hepatocellular carcinoma and metastatic
	CC	liver carcinoma in a patient. The method is useful for identifying
	CC	expression profiles which serve as useful diagnostic markers as well as
	CC	markers that can be used to monitor disease stages, disease progression,
	CC	drug toxicity, drug efficacy and drug metabolism.
	CC	Note: The sequence data for this patent did not form part of the printed
	CC	specification, but was obtained in electronic format directly from WIPO
	CC	at ftp.wipo.int/pub/published_pct_sequences.
	XX	
SQ	Sequence 3084 BP; 859 A; 690 C; 699 G; 836 T; 0 other:	
	Alignment Scores:	
	Pred. No.: 0.119 Length: 3084	
	Score: 98.50 Matches: 52	
	Percent Similarity: 40.11% Conservative: 27	
	Best Local Similarity: 26.53% Mismatches: 82	
	Query Match: 9.44% Indels: 36	
DB:	24 Gaps: 8	
US-10-069-544-2 (1-204) x ABN96929 (1-3084)		
Oy	17 LeuSerAlaCyGSInSeRThPrRoLeProFroLySaEnAmPProGInLeuAlaGlInIle	36
Dd	843 CTCtTCcGTCTGTcT-CcCAccCAGcAGAcTCTCGAATTAC-----CTAGCCAATATC	895
Oy	37 ArgThrGlnIleValIeLSeRLeuEuaSpMetGlyLyLeuAspGlnAlaLysGlnGln	56
Dd	896 AAACGA-----:::GACAGGGAAACATTGAAGAAGCGCTTGCGTTG	934
Oy	57 LeuAspAlaValLeuSerAlaAspArgGlnPhaAlaProAlaYrARGTrHLeuAlaLys	76
Dd	935 TATCGTAAGCATTTNGAAGCTCTCCAGAGTTTGCTGCTGCCATTCAAATTTAGCAAGT	994
Oy	77 ValTyrgInalAserGluAspAlaThrHisgInThryLysAlaGlnIarGlu-----	93
Dd	995 GTACGCAG-----CAGCAGGGGAAAATCGCAGAGCTCTGATGCAT	1036
Oy	94 PheGluTyValAlaIleGluLeuAsnProLySaSpmetGInseTyrmEtAspTyGlyPhe	113
Dd	1037 TATAAGAGCGCTATTCCATCAATCAATCTTCTGCTATAGCCACTCTTAATATAGGAAC	1096

Oy		114	TyrLeuValGlyImeMeGlyAspLeuSerGlyAlaLeuIleTyrPhaAspLysProSerArg	133
Dd		1097	ACTCTAAAGCAATGCACAGATGTTTCAGGAGCCCTTGCACTGTATT-----ACGGCT	114
Oy		134	AlaIleGlyTyrGlUGlyArgValVal---AlaIleGluAsnMetAlaTyrIleTyr---	151
Dd		1148	GCCATCCCAATTAATCCTGCATTTGCAGATGCACATAGCAATCTGGCTTCCATTCAATAAG	120
Oy		152	-----TYRHisGlnTyrGluAlaIalylser	160
Dd		1208	GATTCAAGGAATATTCGAAAGCCATACCTTCTTACCGCACGCCCTCTGAACCTTAAGCCT	126
Oy		161	ProthrLysAspAPryTyraAsnAlaLylsserAlaLeuGluArgAlaLeuIleSerGly	180
Dd		1268	GATTTTCCTGATGCTTATTGTAACTTGCTGCTACTTCCCTGCAG--ATTCTCTGTGATTGG	132
Oy		181	ThrgInHisAapgluilelalyLysSerTyrAspLysLeuLeuSerAsp	196
Dd		1325	ACAGACTATGATGAGCGAATGAGAAGTTGGTCAGTATGTGGCTGCAC	1372
RESULT 28				
AAC98986				
ID	AAC98986	standard; cDNA;	4799 BP.	
XX				
AC	AAC98986;			
XX				
DT	09-MAR-2001	(first entry)		
DE		Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:214.		
XX				
KW	Human; pancreas; pancreatic cancer; pancreatic cancer antigen;			
KW	detection; diagnosis; identification; cytostatic; neuroprotective;			
KV	nocitropic; immunomodulatory; relaxant; contraceptive; gynaecological;			
KM	antiinflammatory; cardiac; gene therapy; chromosome mapping;			
KV	linkeage analysis; tissue identification; tissue typing; forensic;			
KW	neural; immune system; muscular; reproductive; gastrointestinal;			
KV	pulmonary; cardiovascular; renal; proliferative; ss.			
XX				
OS	Homo sapiens.			
XX				
FN	WO200055320-A1.			
XX				
PD	21-SEP-2000.			
XX				
PE	08-MAR-2000; 2000WO-US05989.			
XX				
PR	12-MAR-1999; 99US-0124270.			
XX				
PA	(HUMA-) HUMAN GENOME SCI INC.			
PI	Rosen CA, Ruben SM;			
DR	WPI; 2000-579444/54.			
XX	P-PsDB; AAB54221.			
PT	New nucleic acid that is a pancreatic cancer antigen for preventing,			
PT	treating, or ameliorating a medical condition, particular pancreatic			
PT	cancer, or for use in assays for diagnosing a pathological condition -			
PS	Claim 1; Page 659-660; 1379pp; English.			
XX				
XX	AAC98773 to AAC99231 encode the human pancreatic cancer associated			
CC	proteins, called pancreatic cancer antigens, given in AAB54008 to			
CC	AAB54466. The human pancreatic cancer antigens have cytostatic,			
CC	neuroprotective, nocitropic, immunomodulatory, relaxant, contraceptive,			
CC	gynaecological, cardiac and antiinflammatory activities, and can be used			
CC	in gene therapy. The polynucleotide and proteins can be used for			
CC	preventing, treating, or ameliorating a medical condition or in assays			
CC	for diagnosing a pathological condition or a susceptibility to one in a			
CC	subject. Binding partners to the proteins and the activity of the			
CC	proteins can be identified. The pancreatic cancer antigens can be used to			
CC	detect, treat or prevent pancreatic disorders, especially cancer.			

CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 4799 BP; 1360 A; 947 C; 1005 G; 1478 T; 9 other;

Alignment Scores:

Pred. No.:	0.219	Length:	4799
Score:	98.50	Matches:	52
Percent Similarity:	40.31%	Conservative:	27
Best Local Similarity:	26.53%	Mismatches:	82
Query Match:	9.44%	Indels:	36
DB:	21	Gaps:	8

US-10-069-544-2 (1-204) x AAC98986 (1-4799)

QY 17 LeuSerAlaGysGlnSerThrProIleProProlysaAsnProGlnLeuAlaGlnIle 36
 DB 470 CTCCTCGCTGTGT-CCACCCCATGCAGACTCTCTGATATAC-----CTAGCAATATC 522
 QY 37 ArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGln 56
 DB 523 AAGCA-----GAAAGGAAACATGTAAGAGGAGGAGTTCGCTG 561
 QY 57 LeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyraGlnIleAlaLys 76
 DB 562 TATCGTAAGCATTTAGAGTCTCCAGAGTTGCTCTCCATTCATTAATTAAGCAAGT 621
 QY 77 ValTyrglnAlaSerGlnuAspAlaThrHisGlnThrLysAlaGlnArgLeu----- 93
 DB 622 GTACTGAG-----CAGCAGGAAACTGCGAAGAGCTCGATGCAT 663
 QY 94 PheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrmcAspTyrglyPhe 113
 DB 664 TATAGAGGCTATTCGATCATGCTACCTTCTGATGCTATCTAATATAGGAAAC 723
 QY 114 TyrluValGlnmetGlyAspLeuSerGlyAlaLeuIleTyrlPheAspLysProSerArg 133
 DB 724 ACTTAAAGAGATGAGATGTTCAAGGAGCCTTTCAGAGTGTAT-----ACGCGT 774
 QY 134 AlaIleGlyTyrglnGlyArgValVal---AlaIleGlnuAsnMetAlaTyrlleTyr--- 151
 DB 775 GCCATCCAAATTAATCTCGATTTGCAATGCATATGCAATCTGCTTCATTCATTAAG 834
 QY 152 -----TyrlHisGlnTyrglnAlaAlaLysSer 160
 DB 835 GATTCAGGAATATTCAGAAAGCATAGCTTCTTAACGCGACGGCTGAAACTTAAGCT 894
 QY 161 ProThrLysAspTyrrAsnAsnAlaLysSerAlaLeuGlnuAlaLeuIleSerGly 180
 DB 895 GATTTCCTGATCTTATGTTGTAAGTTCGCTCATTCGCTGAG---ATTGCTGTGATTGG 951
 QY 181 ThrGlnHisAspGlnIleLysLysSerTyrrAspLysLeuLeuSerAsp 196
 DB 952 ACAGACTATGATGAGCGAATGAAGAAGTGTGCTCATGTGTGCTGAC 999
 RESULT 29
 AAX20515/c
 ID AAX20515 standard; DNA; 32768 BP.
 XX AAX20515;
 AC
 XX 05-MAY-1999 (first entry)
 DT
 XX Polynucleotide sequence from the genome of Treponema pallidum.
 DE

XX Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 KW

OS Treponema pallidum.

PN MO9859034-A2.

PD 30-DEC-1998.

PF 23-JUN-1998; 98WO-US13041.

PR 24-JUN-1997; 97US-0050667.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Fraser CM;

DR WPI; 1999-081273/07.

PT New isolated Treponema pallidum nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. pallidum infections, particularly syphilis

PS Claim 1; Page 261-279; 1150pp; English.

CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.

XX Sequence 32768 BP; 8253 A; 9783 C; 7257 G; 7433 T; 42 other;

Alignment Scores:

Pred. No.:	5.42	Length:	32768
Score:	96.50 <td>Matches:</td> <td>36 </td>	Matches:	36
Percent Similarity:	42.19% <td>Conservative:</td> <td>18 </td>	Conservative:	18
Best Local Similarity:	28.12% <td>Mismatches:</td> <td>55 </td>	Mismatches:	55
Query Match:	9.25% <td>Indels:</td> <td>19 </td>	Indels:	19
DB:	20 <td>Gaps:</td> <td>5 </td>	Gaps:	5

US-10-069-544-2 (1-204) x AAX20515 (1-32768)

QY 48 GlyLysLeuAspGlnAlaLysGlnIleuAspAlaAlaLeuSerAlaAspArgGlnPhe 67
 DB 2354 GGAAGATGCAAGAGGAGGACAGACGTAACGACAGTACTG-----CAGGAT 23507
 QY 68 AlaProAlaTyrr---ArgThrLeuAlaLysValTyrglnAlaSerGlnuAspAlaThrHis 86
 DB 23506 ACGCTGACATGAGCGCCGCTTCGACAAATGTGCGCGGTGAAGCGGAACGAACGA 23447
 QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGln 106
 DB 23446 GTGCTGCAAGCATTTAGGATCTGGAACGCTGTGGCGACGCTTGCATCCGATATCTGCA 23387
 QY 107 SerTyrmcAspTyrglyPheTyrluValGlnmetGlyAspLeuSerGlyAlaLeuIle 126
 DB 23386 TATTGACAGATCTCGGACGATTTGTCACAGCGGGA----- 23348
 QY 127 TyrlPheAspLysProSerArgAlaIleGlyTyrglnGlyArgValValAlaIleGln 146
 DB 23347 -----AAAAGAGCGTGGCGAGCGGCTTGCAGCGGCGCATTCCTCTGCT 23297
 QY 147 MetAlaTyrlleTyrrHisGlnTyrglnAlaAlaLysSerProThrLysAspTyrr 166
 DB 23296 GATGCGTAT---TTTGCCCAACATTTATCTGGGGGA-----ATTACAGACAGCTC 23249
 QY 167 AsnAsnAlaLysSerAlaLeuGln 174
 DB 23248 GGGCGAGCGGAGAAAGCATTAG 23225

XX	ABL22195 standard; DNA; 1461 BP.
ID	ABL22195
AC	ABL22195;
DT	26-MAR-2002 (first entry)
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 18058.
KW	Drosophila; developmental biology; cell signalling; insecticide;
RN	pharmaceutical; gene; ds.
OS	Drosophila melanogaster.
MO	MO200171042-A2.
PD	27-SEP-2001.
PP	23-MAR-2001; 2001MO-US09231.
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
PA	(PEKE) PE CORP NY.
P1	Venter JC, Adams M, Li PWD, Myers EW;
DR	WPI; 2001-6556860/75.
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PS	Claim 1; SEQ ID NO 18058; 21bp + Sequence Listing; English.
XX	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-ABBS72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
SQ	Sequence 1461 BP; 381 A; 373 C; 379 G; 328 T; 0 other;

Alignment Scores:	Pred. NO.:	Score:	Length:	Matches:
Percent Similarity:		96.00		39
Best Local Similarity:		45.27%		28
Query Match:		26.35%		63
		9.20%		18
DB:		23	Gaps:	5

US-10-069-544-2 (1-204) X ABL22195 (1-1461)

Oy	64 AsparGlnInpheAlaPheAlaTyrAgtgTrtLeuAlaIysValATyrGlnAlaSerGIuAsp	83
Db	181 AATCGCCATTGGATTCAGAGCCTGTACTTTGCCAGGCCTTAATTGACCGGAAGAG	240
Oy	84 AlaThrTgInthrLysAlaGlnArgLeuPheGluYlValAlaIleGluLeuAsnProIvs	103
Db	241 GGCAACAC--ATCGAGGCGCTCAACAATTGGCAAATAATCGCGAGCTTAATCCAGG	297
Oy	104 AspMetGlnSerTyrrMetAspTYrGlyPheTyrlEuvalGlnMetGlyAspLeuSerGly	123
Db	298 AATATCGACATTACAAAGAAATCGGCCAACACTATACATATGCGCGTTTAGTGAG	357
Oy	124 AlalauleIeTyrrheaaplysrProserArGaIlaIleGlyTYrGluGlyArGValValaIa	143

Db 358 GCTTTGGGCGTTTTCGGAGCGC-----GAACAAAGCAGACGTCCA 399
 Qy 144 ILeGuaaMeaLaTyrlLeTyrlRhIsGln-----TyrgIuaIaIaIySeSerPro 161
 Db 400 CAGAGACCAAGATTTACCATTAATCTGTGGCGAGCACTCTATAGAGCGGCACAGCGCAG 459
 Qy 162 ThLyAspAsp-----TyrlaMbnaIaIySeSerAlaLeuGIuaYgAla 176
 Db 460 AGCAAAAGAAATGTGGCCAGCCAGCAGACGAGAGACTATTTTGTAGTTGGCC 519
 Qy 177 LeuIleSeGjIYhrgIhIaSpGIuIleIySlySeSerTyraBpLySleuLeuSerAsp 196
 Db 520 GTGCAATCCGCGCCGAG-----CTGAAAGACTTACGTTCCGTTGGCGAGCTTG 567
 Qy 197 TyrlYsleuLeuSerAspTyrlYls 204
 Db 568 TATCGCAAGACAAACAGTACAG 591
 RESULT 31
 ID AAX99595 standard; DNA; 2166 BP.
 AC AAX99595;
 AC AAX99595;
 DT 05-OCT-1999 (first entry)
 XX
 XX Nucleic acid sequence from U. urealyticum.
 DE
 XX Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;
 KW human urogenital tract; pregnancy; neonatal disease; drug therapy;
 KM suppurative arthritis; ss.
 XX
 XX Ureaplasma urealyticum.
 OS
 PN WO9393007-A1.
 XX
 XX 05-AUG-1999.
 PD
 PF 29-JAN-1999; 99WO-US01972.
 XX
 PR 30-JAN-1998; 98US-0073189.
 XX
 XX (UABR-) UAB RES FOUND.
 PA
 XX Casell GH, Chen EY, Glaes JI, Glaes JS, Heiner CR;
 PI Leckowitz E;
 PI WPI; 1999-469343/39.
 DR
 XX
 XX Detection of Ureaplasma urealyticum using novel genes, probes and
 PT primers
 PS
 XX Claim 1, Page 69-70; 110pp; English.
 CC The present invention provides methods for the detection and diagnosis
 CC of Ureaplasma urealyticum infection. It provides novel genes
 CC (AAX99501-661) that can be used as a source of primers and probes for the
 CC detection and/or quantification of U. urealyticum in a biological
 CC sample. The probe that can be used in the method of the invention by
 CC forming target:probe complex is complementary to a region selected from
 CC one of the 181 nucleic acid sequences (AAX99501-661). U. urealyticum is
 CC an opportunistic pathogen of the human urogenital tract that is a
 CC significant cause of adverse pregnancy outcome, neonatal disease, and
 CC suppurative arthritis. As the infections are commonly asymptomatic, it is
 CC important to have specific and sensitive methods for detecting their
 CC presence in a patient. Also, as the pathogen has no current antibiotic
 CC directed specifically against it, it would be advantageous to isolate and
 CC detect gene sequences which are unique to it, and utilize these as a
 CC basis for diagnosis of U. urealyticum infection as well as to develop new
 CC and improved drug therapies. The present invention provides such novel
 CC polynucleotide sequences (AAX99501-661).
 XX

SQ Sequence 2166 BP; 940 A; 252 C; 265 G; 709 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.172 Length: 2166
 Score: 95.50 Matches: 56
 Percent Similarity: 35.12% Conservative: 29
 Best Local Similarity: 23.14% Mismatches: 76
 Query Match: 9.16% Indels: 81
 DB: 20 Gaps: 13
 US-10-069-544-2 (1-204) x AAK9595 (1-2166)
 QY 1 MetLysIleArgValYsTrp-----ProMetValMetAlaMet 13
 13 TTGAAATATAAATTAATTTAGCTTCACTAGCAACAATCAATGTA---GCTTCA 69
 QY 14 GlyLeuValLeuSerIaCys----- 20
 70 CCATTATATTTTCTGTGACGCTAATGAGATGTAAACAAACAAATATGCTAAT 129
 QY 21 -----GlnSerThrProIleProProLysAsnAsnProGlnLeuAla 34
 130 AATCATGATCTTAATCAACAAAAAACCAAAATTCCTAAACGATCTAATTTAAT 189
 QY 35 GlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLys 54
 190 ACTTTTAAACT-----GATCCTGATTA---ACTGTTAA 222
 QY 55 GlnGlnLeuAspAlaIleLeuSerAlaAspArgGlnIleAlaProAlaTyrArgThrLeu 74
 223 GATAGTTTAGAAAAAGAAATTAATGCT----- 249
 QY 75 AlaIysValTyrGlnAlaSerGluAspAlaThrIleGlnThrLysAlaGlnArgLeuPhe 94
 250 GCATTTGTAATGTAATCAACGTCAGAAATAATTGAATAAATAAATAATGAGTTT 309
 QY 95 GlnLysAlaIleGlu-----LeuAsnProLysAspMetGlnSerTyrMet----- 109
 310 AAGAAAAAATTAACAACTAATTTATTTAAAAATCTGACATCATCTTGAGAAAAAT 369
 QY 110 -----AspTyrGlyPheTyrLeuValGlnMetGlyAsp 120
 370 AAAGAAACATCTTAATAAATCCTAATGATTAATGTTTATTTA---AATACCTCCAA 426
 QY 121 LeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArg 140
 427 ATTTAGCAACTCTTAATAAATTAATGAT-----ATTAAAGATATTTGATTAATGAGAA 480
 QY 141 ValValAlaIleGluAsnMetAlaTyrIleTyrThrIleGlnTyrGluAlaIleLysSer 160
 481 ACA-----TATAAGCAATCAAAAGTTGGCAAACT 510
 QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
 511 -----GATCTTTAATTAATCAAAAGCCCTGCTCCAAAAGAAAAAATTAATGAT 561
 QY 181 ThrGlnHisAspGluIle-----LysLysSerTyrAspLysLeuLeuSerAsp 196
 562 GTACAAAGTATCAATTAATGATGTGAGAAACAAATAACAAAGATACGTTAAAAAAA 621
 QY 197 TyrLys 198
 622 TATGAA 627
 RESULT 32
 ABR87389
 ID ABR87389 standard; DNA; 1758 BP.
 AC ABR87389;
 XX
 DT 24-SEP-2002 (first entry)
 XX
 DE Target molecule human PAT1 cDNA.

XX Retinoic acid pathway; RA; ss; gene; retinoid; lung cancer;
 KW Kaposi's sarcoma; breast cancer; pancreatic cancer; neuroblastoma;
 KW renal cancer; ovarian cancer; dermatitis; hyperkeratosis; eczema;
 KW Darier's disease; Reiter's disease; psoriasis; gene therapy; R2;
 KW acute promyelocytic leukaemia; APL; Perturbagen; R3; F802; F820;
 KW yeast two-hybrid assay; PAT1; kinesin light chain-related protein;
 KW human.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..1758
 FT /tag= a
 FT /product= "PAT1"
 PN WO200240719-A2.
 XX
 PD 23-MAY-2002.
 XX
 PF 17-NOV-2001; 2001WO-US44039.
 XX
 PR 17-NOV-2000; 2000US-249468P.
 XX
 PA (DELT-) DELTAGEN PROTEOMICS INC.
 XX
 PI Kamb CA, Richards BT, Karpilow J;
 XX
 DR WPI; 2002-519386/55.
 DR P-PSDB; AAU99188.
 XX
 PT Polypeptide with retinoic acid pathway activity, especially of
 PT perturbagens R3, F802 and F820 for identifying a cellular target which
 PT interacts with the polypeptide and for therapeutic purposes
 XX
 PS Claim 104; Fig 16c; 131pp; English.
 CC The invention relates to an isolated polypeptide (I) with retinoic acid
 CC (RA) pathway activity, comprising a polypeptide sequence of Perturbagen
 CC (phenotypic probe) R3, Perturbagen F802, Perturbagen F820, their
 CC biologically active modifications, or biologically active fragments.
 CC Also included are the polynucleotides encoding the perturbagens, a
 CC vector comprising the polynucleotide, preparing an RA pathway related
 CC polypeptide, a composition comprising the polypeptide, an antibody to the
 CC polypeptide, screening (M) putative RA-related therapeutics, by exposing
 CC a polypeptide/target interaction pair identified by a method utilizing
 CC the RA pathway polypeptide, to a number of agents, and recovering a
 CC subpopulation of disrupting agents which competitively displace the
 CC polypeptide from the target, where the disrupting agents are putative RA-
 CC related therapeutics. Also include are an isolated RA pathway polypeptide
 CC comprising PAT1 (a kinesin light chain-related protein) polypeptide
 CC and its encoding polynucleotide, a gene therapy vector comprising the RA
 CC pathway protein polypeptide or encoding or PAT1 and a host cell
 CC comprising the gene therapy vector. The RA pathway polypeptide is useful
 CC for identifying a cellular target that interacts with RA pathway-related
 CC polypeptide, by exposing the polypeptide in vitro to putative target
 CC molecules and identifying a polypeptide/target interaction pair, by
 CC detecting reporter expression, where the reporter expression is
 CC operatively linked to the formation of the interaction pair (M) is a
 CC yeast two-hybrid assay. The polypeptide is also useful for treating
 CC an RA pathway-related condition e.g. lung cancer, Kaposi's sarcoma,
 CC breast cancer, pancreatic cancer, neuroblastoma, renal cancer,
 CC ovarian cancer, dermatitis, hyperkeratosis, eczema, Darier's disease,
 CC Reiter's disease, psoriasis, acute promyelocytic leukaemia (APL).
 CC The present sequence is a cDNA encoding a target
 CC molecule isolated by the yeast two hybrid system, PAT1.
 CC
 XX
 SQ Sequence 1758 BP; 526 A; 330 C; 391 G; 511 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.149 Length: 1758
 Score: 95.00 Matches: 45
 Percent Similarity: 39.43% Conservative: 24

Query Match: 9.11% Indels: 50
DB: 24 Gaps: 8
US-10-069-544-2 (1-204) x ABL64883 (1-6465)

QY 47 MetGlyLysLeuSerGlnAlaLysGlnGln-----Leu 57
DB 865 ATGGATAAATCATATCAAAACATGCGCAGCAACAAATAAGCTGCATCTATGAGAACTG 924
QY 58 AspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArg-----72
DB 925 TGTGACTCTCTATTGCAAAAAGTCACTATGATGAGCAATACAAATGTCATCGACGCA 984
QY 73 -----ThreAlaAlaLysValTyrGlnAlaSerGlnAspAlaThrHisGlnThr 88
DB 985 ATGAAGAATAATACAGACGGCTTACAGTGAAGTTGTGTGGATGTCTTAAGACAAAGCT 1044
QY 89 -----LysAlaGlnArgLeuPheGluLysAla 97
DB 1045 TCTAAGGCTTGTGTAGTAAACGTGAATTAAGAAAGCAGACAGTAAATTAACATGCA 1104
QY 98 IleGluLeu-----AsnProLysAspMetGlnSerTyrMet 109
DB 1105 GTGTATTGGCAGCGGATCATTTTGGATCCAAACCCAAATAATCTGATACATGCTTA 1164
QY 110 AspTyrGlyPheTyrLeuValGlnMetGlyAspLeu-----SerGlyAlaLeuIleTyr 127
DB 1165 GATTATGGTCTTACTTACTCATGTATGATATATCTGTGATGCTGTTCGATTTATCAG 1224
QY 128 PheAspLysProSerArgAlaIleGlyTyrGlnGlyArg-----ValValAla 143
DB 1225 GCAGCCCTTGACATTAGACAGTCACTGTGTGGCGCAAAATATCCAGTACGACAGCT 1284
QY 144 IleGluAsnMetAlaTyr---IleTyrTyrHisGlnTyrGluAlaAlaLysSerProThr 162
DB 1285 CATGAAGATTGGCTCACTCTTCTTATGTCACACAGATATGCTCTGGGAAA-----1335
QY 163 LysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeu 177
DB 1336 -----TTTGACAAATGCACTATTTCATGACGAAAGAGCTATT 1371

RESULT 34
AAC47531
ID AAC47531 standard; DNA; 1246 BP.
XX
AC AAC47531;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54164.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 22-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 03-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 0.163 Length: 1246
Score: 93.00 Matches: 51
Percent Similarity: 38.58% Conservative: 25
Best Local Similarity: 25.89% Mismatches: 87
Query Match: 8.92% Indels: 34
DB: 21 Gaps: 7

US-10-069-544-2 (1-204) x AAC47531 (1-1246)
QY 12 AlawetgyleuvalleuSerAlCysglInserThrProIleProProlysaAsnaPro 31
DB 254 GCAGCAGCAGCTGTCCTGTAATCTCCAGCTCAACCTCAGCGGCGGCTCCT 313
QY 32 GlnleuAlaglnleuThrcInleAlaleSerleuAmpMerGly----- 48
DB 314 GTGGGGGG-----ACGCCGTATATGGAATCTTCACAAACAAAGTAACGAAACGTT 364
QY 49 LysleuAepglnAlaLysGlnleuApsAlaAlaLeuSerAlaAspArgGlnPheAla 68
DB 365 TCATTGGAAGAAGAGAGATCACTCGAAGAGATCTGCTTCACTCACTGATGAT 424
QY 69 ProAlaTyArgThrleuAlaLysValTyGlnAlaSerGluAspAlaThrHisGlnThr 88
DB 425 GAAGCTACGATCTTTAATGAGGTAAAGATCAATCTGAAAGCTT-----CTA 475
QY 89 LysAlaGlnArgleuPheGlnLysAlaAlleuLeuAenProLysAspMetGlnSerTy 108
DB 476 GAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 535
QY 109 MetAspTyArgPheTyLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyPhe 128
DB 536 ATGTTGAAAGCTAATATCTTATGATCACTGAGATTTAGAGTCTGCTTAAACTGTTT 595
QY 129 Asplys-----ProSerArgAlaIleGlyTyArgGlyArgValAlaAla 143
DB 596 GAAGAGATTTAGTCAAGATCCGCTTCGTCGAGGCTTACCGGTTAGTT----- 649
QY 144 IleGluAenMetAlaTyIleTyTyHisGlnTyArgLysAlaLysSerProThrLys 163
DB 650 -----ATGGCTTAT-----TCGATTCGCT 670
QY 164 AspAspTyraenAenAlaLysSerAlaLeuGluArgAlaLeuIleSerGlyThrGlnHis 183
DB 671 GATGATTTGATGCTGTCGGAAGAAGATTAAGAGCAATGATGATGATGATGATGAT 724
QY 184 AspGluIleLysLysSerTyraAspLysLeuLeuSerAspTyTyLysLeuLeu 200

PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149829.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	22-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Alignment Scores:	
Pred. No.:	0.163
Score:	93.00
Percent Similarity:	38.58
Best Local Similarity:	25.89
Query Match:	8.92
DB:	21
US-10-069-544-2 (1-204) x AAC40535 (1-1248)	
	Length: 1248
	Matches: 51
	Conservative: 25
	Mismatches: 87
	Indels: 34
	Gaps: 7

Dy 12 A l a n e t g l y l e u v a l l e u s e r a l c y s g l i n s e r t h r p r o l i e p r o p r o l y s a s n a s n d r o 31
 ||| ||| ||| ||| |||
Db 256 G C A C A G A C T G T T C C T T G A A T C C A A C C T T C A C C G G A T T G C G C T C C T 315

Oy	32	GlnLeuAlaGlnIleArgThrcInIleAlaIleSerLeuLeuAspMetGly-----	48
Db	316	GTGGCGGCG-----ACGCCCTTAATGGAATCTCTCAACCAAGTAACGAAACGTT	366
Oy	49	LyLeuAspGlnIleAlaGlnGlnIleuAspAlaIleuSerAlaAspArgGlnPheAla	68
Db	367	TCATTGGAAGAAGAAAGAGATCACTGAAAGATATCTTGGTCTTCAACCTGATGATGT	426
Oy	69	ProIlaTyrArgThrcLeuAlaIaIyValTyrGlnIlaSerGlnuAspAlaIthrIhsglnThr	88
Db	427	GAAGCTTACGATCTCTTATATGAGAGGTATGATCAAACTCTGAAAGCTT-----CTA	477
Oy	89	LySaIaGlnArgLeuPheGlnIuValaIleGlnIleuAsnProLyAspMetGlnSerTyr	108
Db	478	GAAGCTATATGAATTGATTGATGCTTGCTTGAATTTGGAACCAAGGAAAGAAATGCCCA	537
Oy	109	MetAspTyrGlyPheTyrIleuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPhe	128
Db	538	ATGTTGAAGCTAAATATCTTAATCTTAGTACAGTGGAGCTTTAGAGCTGCTTAAACGCTTT	597
Oy	129	AspLys-----ProSerArgAlaIleGlyTyrGlnGlyArgValaIa	143
Db	598	GAAGAGATTCTAGTCAAGATCCGCTTGTGTGCGAGGCTTATCAGGCTTTAATGTT-----	651
Oy	144	IleGluAsnMetAlaTyrTyrIleTyrThrIhsglnTyrGlnIlaIaIaLysSerProThrLys	163
Db	652	-----ATGCTTAT-----TCGATTCGGT	672
Oy	164	AspAspTyrAsnAsnAlaLysSerAlaLeuGlnuArgAlaLeuIleSerGlyThrGlnIh	183
Db	673	GATGATTGAAATGCTGTGAGACAGAGATTAAGAGGAAATGCTTAAGTGCAG-----	726
Oy	184	AspGlnIleLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeu	200
Db	727	-----AAGAGAGAAGATCGTAAAGACTTGAAGATTTAAGCTACTA	768

RESULT 36

AAFP3894	standard, cDNA, 3080 BP.
ID	AAFP3894
XX	AAFP3894;
XX	
DT	23-MAY-2001 (first entry)
DE	Human cDNA encoding a membrane or secretory protein clone PSEC0261.
KM	Human; secretory protein; membrane protein; vaccine; gene therapy;
KW	rheumatoid arthritis; diabetes; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1067182-A2.
XX	
PD	10-JAN-2001.
XX	
PF	07-JUL-2000; 2000EP-0114090.
XX	
PR	08-JUL-1999; 99JP-0194179.
PR	11-JAN-2000; 2000JP-0118775.
FR	02-MAY-2000; 2000JP-0183766.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Oca T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX	
DR	WPI; 2001-093989/11.
XX	
XX	P-PSDB; AAB88467.
PT	Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX	gene therapy or as candidate target molecules in drug development -
XX	
XX	Claim 1; SEQ ID 301; 609pp + CD ROM; English.

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 SQ Sequence 3080 BP; 907 A; 658 C; 703 G; 812 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 0.651 Length: 3080
 Score: 92.50 Matches: 35
 Percent Similarity: 44.53% Conservative: 26
 Best Local Similarity: 25.55% Mismatches: 69
 Query Match: 8.87% Indels: 7
 DB: 22 Gaps: 4
 US-10-069-544-2 (1-204) x AAF93894 (1-3080)
 QY 28 LysAsnAnpProGlnLeuAlaGlnIleArgThrglnIleAlaIleSerLeuLeuAspMet 47
 DB 700 AAAAACCCAGCCT--ATAGCTATGCTATACAA-----GGTTTAACTTCTTTCACAGA 750
 QY 48 GlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaLeuSerAlaAspArgGlnPhe 67
 DB 751 GGACTTCTGAAGAGACTTATGATCTCTCAAGAACCTTTGAAGACAGAAAGTTGACTTT 810
 QY 68 AlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAspAlaThrHisGln 87
 DB 811 ATTGATGCAATATAAAGTCTAGGAGCATATAGAACTGGGCAATTGAGACAGACC 870
 QY 88 ThrLysAlaGlnArgLeuGlnGlnLysAlaIleGlnLeuAspProLysAspMetGlnSer 107
 DB 871 ACTGAGAGC-----TTTCAAAAGGCACTGTTGCTCAACCAAAATCATGTGCAAAACC 921
 QY 108 TyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyr 127
 DB 922 CTCACAGCTCCGGGAGATGATGCTCTACACACAGCAGCTTACAGAAACCTTAAGAAC 981
 QY 128 PheAspLysProSerArgAlaIleGlyTyrGlnGlyArgValAlaIleGlnAsnMet 147
 DB 982 TTTAAGCGGTGCTGACGCTAGAGCATATATGAAGTGTGACAGATATATGAAGGCGTTC 1041
 QY 148 AlaTyrIleTyrTyrHisGln--TyrGlnAlaAlaLysSerProThrLys 163
 DB 1042 ACCCATGTTGCCATGGACAGTTTATGAAAGGATAAAGCAACAAACAAA 1092

DE Human cDNA encoding a membrane or secretory protein clone PSEC0076.
 XX
 XX Human, secretory protein; membrane protein; vaccine; gene therapy;
 KW Rheumatoid arthritis; diabetes; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1067182-A2.
 XX
 PD 10-JAN-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114090.
 XX
 PR 08-JUN-1999; 99JP-0194179.
 PR 11-JAN-2000; 2000JP-0118775.
 PR 02-MAY-2000; 2000JP-0183766.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isegai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX
 DR WPI, 2001-093989/11.
 XX
 P P-PSDB; AAB88351.
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 XX
 PS Claim 1; SEQ ID 69; 609pp + CD ROM; English.
 XX
 This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 SQ Sequence 3253 BP; 956 A; 687 C; 754 G; 856 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 0.702 Length: 3253
 Score: 92.50 Matches: 35
 Percent Similarity: 44.53% Conservative: 26
 Best Local Similarity: 25.55% Mismatches: 69
 Query Match: 8.87% Indels: 7
 DB: 22 Gaps: 4
 US-10-069-544-2 (1-204) x AAF93778 (1-3253)
 QY 28 LysAsnAnpProGlnLeuAlaGlnIleArgThrglnIleAlaIleSerLeuLeuAspMet 47
 DB 846 AAAAACCCAGCCT--ATAGCTATGCTATACAA-----GGTTTAACTTCTTTCACAGA 896
 QY 48 GlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaLeuSerAlaAspArgGlnPhe 67
 DB 897 GGACTTCTGAAGAGACTTATGATCTCTCAAGAACCTTTGAAGACAGAAAGTTGACTTT 956

```

Qy      68 AAlaProAieTYrargThrlLeuaLalysValTYrglInlaIsenGluAspaLaThrHleGlN    87
           |||||.....:::|||||...:::
Db       957 ATTGATGATATAAAAGCTTAGGGCAGGCATTATACGAACCTGGCAATTTTGAMGACC    1014
Qy       88 ThrLySaIacInAgLeuPheglulYsaIaiIeguleAasProLYsaSPmetGIser    107
           |||:::|||||.....:::|||||...:::
Db       1017 ACTGAGAGC-----TTTCAAAGGCACTGTTCCTCAACCAAAATCATGTGCCAAC    106
Qy        108 TyrmelaePrTyRgIphneTYreUValelmecGLYaSpIusecILYAaleuIEtyR   127
           |Oe8 CTCACAGCTCCGGAGAATGATGCTCTCACCAACAACCGCACTTAACAGAAAGCCCTTAAGAAC 1127
Qy       128 PheaBlyPePseSerARAlaillegITyrGlugIArgVAIlvalAIaliegIUasmEc    147
           |||||.....:::|||||...:::
Db       1128 TTTAAGCCGCTCTCGCACTAGAGCAATTAATGAAGCTGCAGTAAATGAAGGCTC    116
Qy       148 AlAtYrileTYrTYRHleGlN--TYrGUAlaalalySeserProTHryLye     163
           ::::::|.....:::|||||...:::
Db       1188 AGGCATGTTCACATGAGCAGACTTTATGAGAGGATAAAAGCAACAAACAAAA    1238

RESULT 38
ABLO5085
ID      ABLO5085 standard; cDNA; 1979 BP.
XX      ABL05085;
XX      DT
DT       26-MAR-2002 (first entry)
DE      Drosophila melanogaster expressed polynucleotide SEQ ID NO 9737.
XX      XX
XX      Km pharaceutical; gene; ss.
XX      OS
OS      Drosophila melanogaster.
XX      WO200171042-A2.
XX      PD
PD       27-SEP-2001.
PF       23-MAR-2001; 2001WO-US09231.
PR       23-MAR-2000; 2000US-191637P.
PR       11-JUL-2000; 2000US-0614150.
XX      XX
XX      (PEKE ) PE CORP NY.
XX      PA
PA       Venter JC, Adams M, Li PWD, Myers EW;
PI       WPI; 2001-656860/75.
DR       P-PDSB; ABB60982.
PT       New isolated nucleic acid detection reagent for detecting 1000 or more
PT       genes from Drosophila and for elucidating cell signalling and cell-cell
PT       interactions -
PS       Claim 1; SEQ ID NO 9737; 21pp + Sequence Listing; English.
PX      The invention relates to an isolated nucleic acid detection reagent
XX      capable of detecting 1000 or more genes from Drosophila. The invention is
XX      useful in developmental biology and in elucidating cell signaling and
XX      cell-cell interactions in higher eukaryotes for the development of
XX      insecticides, therapeutics and pharmaceutical drugs. The invention
XX      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX      sequences (ABL01840-ABL16175) and the encoded proteins
XX      (ABB57737-ABB72072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pcc_sequences.
XX      Sequence 1979 BP; 600 A; 512 C; 481 G; 386 T; 0 other;

Alignment Scores:
```

Pred. No.:	0.47	Length:	1979
Score:	91.50	Matches:	28
Percent Similarity:	39.02%	Conservative:	20
Best Local Similarity:	22.76%	Mismatches:	57
Query Match:	8.77%	Indels:	19
DB:	23	Gaps:	2

US-10-069-544-2 (1-204) x ABL05085 (1-1979)

```

Oy      4 ArgValIysTPrPro-----8
          |||
Db      1045 AGGAGCGAATRGCTCATCATCCGAGAAAGCCGAGAGAAGACGAGCGCAACC 1104
Oy      9 MetValMetAlaMetGIlyLeuValIleuSerAlaCySgInserThrProIleProPolyS 28
          :|||
Db      1105 TTTTCCTTAGAAGAGGGGACTACAGCACCGCTGTAAAGCACTACACGAGGCCATC-AAG 1163
Oy      29 AsnAsnProGlnIleuAlaGlnIleAlxThrGlnIleAlaIleSerIleuLeuSpMetCgly 48
          |||||
Db      1164 CGCAATCCGGATGATGCCCAACTCTTAATAGTAACCCGCCCTGCCTACACCAAGCTGGCC 1223
Oy      49 LysLeuSpGlnAlaLysGlnGlnLeuAspAlaAlaIleuSerAlaAspArgGlnPheAla 68
          |||||
Db      1224 GCCTTTGATCTGGCGCTCAGAGACTGCAGACCTTGCACTCAAGCTAGACGAATAATTCTATA 1283
Oy      69 ProAlaTYrArGThrIleuAlaLysValTyrglnAlaSerGluAspAlaTrnHisGlnThr 88
          |||||
Db      1284 AAGGAGCTAACATCCGCAAGGCGCAAAATTCATACAGGCGATGCAG-----CAACAATCT 1334
Oy      89 LysAlaGlnArgLysPheGluLysAlaAlaIleGlnLeuAsnProLysAspMetGlnSerTyR 108
          |||||
Db      1335 AAGCGCGCAAGCGCGCTTCCAAAAGGCCCTGTGAATGTGACCCCAACAACGCGAGGCGCAATT 1394
Oy      109 MetAspTYr 111
          |||
Db      1395 GAAGCTTAC 1403

RESULT 39
ABL05084/C
ID     ABL05084 standard; cDNA; 4284 BP.
XX
XX     ABL05084;
AC
XX     26-MAR-2002 (first entry)
DT
XX
XX     Drosophila melanogaster expressed polynucleotide SEQ ID NO 9734.
DE
XX
XX     Drosophila; developmental biology; cell signalling; insecticide;
KW     pharmaceutical; gene; ss.
OS
XX
XX     Drosophila melanogaster.
PN
XX     WO200171042-A2.
PD
XX     27-SEP-2001.
PF
XX     23-MAR-2001; 2001WO-US09231.
PR
XX     23-MAR-2000; 2000US-191637P.
PR     11-JUL-2000; 2000US-0614150.
XX
XX     (PEKE ) PE CORP NY.
PA
XX
XX     Venter JC, Adams M, Li PWD, Myers EW;
PI
XX     WPI; 2001-656860/75.
DR
XX     P-FSDb; ABB60981.
PT
XX     New isolated nucleic acid detection reagent for detecting 1000 or more
PT     genes from Drosophila and for elucidating cell signalling and cell-cell
PT     interactions -
XX
XX     Claim 1; SEQ ID NO 9734; 21bp + Sequence Listing; English.
```



```
Db 853 GCCTATGAGAGAGAGCTGTGTCTCTTCAGATGGGTAAATTTTGTGCATGCAAG 912
Oy 83 AspAlaThr----- 85
Db 913 GATATTATGCTGCCATGAGATCTACTACAGCAGAAATCTTACAAATCGTGGGTG 972
Oy 86 -----HisGlnThrLysAlaGlnArgLeuPheGluLysAlaIleGlu 99
Db 973 ATTATGAGTTTATGGGCACAAACAGATGCAATGAAGAAGCTACCAAGATGCAATTACT 1032
Oy 100 LeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGly 119
Db 1093 CTTAAACCCCAAGTACTCGCTGCTTACTTAAATGCAAGAAATATCTACTTCCACCACAG 1092
Oy 120 AspleuSerGlyAlaLeuIleTyrPheAspLysPro----- 131
Db 1093 CAGTTTCCAGGCCAGCTGACTTCTCAAAAGCTTTAAATTTGATCCAGAAATGAA 1152
Oy 132 -----SerArgAlaIle-----GlyTyr 137
Db 1153 TATGTTCTCATGAATCGAGCTATTACAAATACAAATATTAAAGAAATATGAAGCAAAA 1212
Oy 138 GluGlyArgValValAlaIleGlu-----AsnMet 147
Db 1213 GAAGATTTTGCAAAATGTAATTGAAGCTGTCCCTTTGGGCTGCAGTATATTTTATAGA 1272
Oy 148 AlaTyrIleTyrTyr-----HisGlnTyrGluAlaIleLysSerProThrLysAspAsp 165
Db 1273 GCACATTTTCTACTACTGCTTAAAGCAATATGAACTAGCT-----GAGGAAAGAC 1320
Oy 166 TyrAsnAsnAlaLysSer-----AlaLeuGlu-----ArgAlaLeu 177
Db 1321 CTTAATAAAGCCCTGTCTTTAGAGCTAATGATGCTCTAGATATATATTTTAAAGCAAAA 1380
Oy 178 IleSerGlyThr-----GlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSer 195
Db 1381 GTTCGTGTAAAAATAGTCTGTGATGAGAGCTATGCTGACTATATACCAAGCACTTGAT 1440
Oy 196 AspTyrLysLeuLeuSerAspTyr 203
Db 1441 -----CTTGAAGACTAT 1452
```

Search completed: July 4, 2003, 03:55:52
Job time : 2492 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 8, 2003, 11:21:44 ; Search time 70 Seconds
(without alignments)
388.330 Million cell updates/sec

Title: US-10-069-544-2
Perfect score: 1043
Sequence: 1 MKIRVWPMWMMGLVLSAC.....EIKSYDKLLSDYKLLSDYK 204

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

```

1: A.Geneseq_101002.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1043	100.0	204	22	AA060638
2	124.5	11.9	1151	19	AA062501
3	117.5	11.3	253	21	AA075451
4	116	11.1	1059	22	AB059330
5	116	11.1	1059	22	AB059330
6	116	11.1	1059	22	AB059330
7	115	10.7	253	21	AA075450
8	106.5	10.2	230	23	AB038764
9	106.5	10.2	408	22	AA093717
10	106.5	10.2	981	19	AA046459

11	106	10.2	264	20	AA032160
12	103.5	9.9	253	21	AA075452
13	101.5	9.7	212	23	AB049001
14	99	9.5	230	22	AA032898
15	98.5	9.4	920	19	AA082500
16	96	9.2	486	22	AB067914
17	95	9.1	585	23	AA099188
18	93	8.9	333	21	AA024834
19	93	8.9	333	21	AA043345
20	93	8.9	344	21	AA043344
21	93	8.9	345	21	AA024833
22	92.5	8.9	806	22	AA088467
23	92.5	8.9	860	22	AA088351
24	91.5	8.8	490	22	AB059359
25	91	8.7	387	21	AA079489
26	90.5	8.7	465	21	AA042251
27	90.5	8.7	473	21	AA042250
28	90.5	8.7	482	21	AA042249
29	90	8.6	165	23	AB059759
30	90	8.6	234	21	AA024835
31	90	8.6	234	21	AA043346
32	90	8.6	900	13	AA020568
33	90	8.6	1115	21	AA014972
34	87.5	8.4	1091	16	AA070982
35	87.5	8.4	1091	16	AA070983
36	86	8.2	160	22	AA023157
37	86	8.2	714	18	AA035316
38	86	8.2	1238	22	AB062022
39	85.5	8.2	251	21	AA029689
40	85.5	8.2	265	21	AA029688
41	85.5	8.2	375	21	AA029687
42	85.5	8.2	380	22	AA027645
43	85.5	8.2	519	21	AA079211
44	85.5	8.2	519	22	AA093960
45	85.5	8.2	519	23	AB043363

ALIGNMENTS

```

RESULT 1
AAB060638 standard; Protein; 204 AA.
ID
XX
AC AAB060638;
XX
DT 02-MAY-2001 (first entry)
XX
DE Moraxella catarrhalis strain ATCC43617 BASB121 protein.
XX
DE BASB121 protein; strain ATCC43617; antigen; antibody; vaccine;
XX
KW Genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory.
XX
OS Moraxella catarrhalis.
XX
PN WO200109330-A2.
XX
PD 08-FEB-2001.
XX
PF 27-JUL-2000; 2000WC-EP07281.
XX
PR 30-JUL-1999; 99GB-0018040.
XX
PI (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonard J;
XX
WPI: 2001-159868/16.
XX
N-PSDB; AAF59768.
XX
PT New polypeptides and polynucleotides of Moraxella catarrhalis, useful

```

Soybean CDC-16 pro
Neisseria meningit
Listeria monocytog
Novel human secret
Human OGT protein.
Drosophila melanog
Target molecule hu
Arabidopsis thaliana
Arabidopsis thaliana
Arabidopsis thaliana
Arabidopsis thaliana
Human membrane or
Human membrane or
Drosophila melanog
Human testis-speci
Arabidopsis thaliana
Arabidopsis thaliana
Arabidopsis thaliana
Arabidopsis thaliana
Human OREX protein
Arabidopsis thaliana
Sequence of a part
N. meningitidis an
Megakaryocyte stem
Megakaryocyte stem
M. catarrhalis R1
Drosophila melanog
Arabidopsis thaliana
Arabidopsis thaliana
Human protein APP6
Human transferrase
Human polypeptide,
Human flavoprotein

PT as vaccine for prevention, treatment of microbial diseases and in
PT diagnostic assays for detecting diseases associated with microbial
PT infections

PS Claim 4; Page 66; 81pp; English.

XX The invention relates to the Moraxella catarrhalis strain ATCC43617
CC BASB121 protein (AAB60638) and to DNA encoding it (AAB59768). The
CC invention also relates to immunogenic fragments of the BASB121 protein,
CC expression vectors and host cells comprising BASB121 nucleic acids, the
CC recombinant production of BASB121, vaccine compositions comprising the
CC BASB121 protein or nucleic acid, an antibody against BASB121, therapeutic
CC compositions comprising the anti-BASB121 antibody, and a method of
CC identifying a Moraxella catarrhalis infection via the detection of
CC BASB121 proteins or antibodies. The vaccine compositions of the invention
CC are useful as prophylactic or therapeutic agents against Moraxella
CC catarrhalis infections in mammals, particularly humans. Moraxella
CC catarrhalis is a Gram negative bacterium frequently isolated from the
CC human upper respiratory tract, which is responsible for several
CC pathological conditions. It is responsible for about 15% of otitis media
CC cases in children (which can lead to temporary or permanent hearing
CC loss). It also causes pneumonia in elderly people, and sinusitis,
CC nosocomial infections and, less frequently, invasive diseases. BASB121
CC proteins or nucleotides may additionally be used in screening for novel
CC antibacterial compounds, and in the diagnosis and staging of infections.
CC The present sequence represents the Moraxella catarrhalis strain
CC ATCC43617 BASB121 protein.

XX Sequence 204 AA;

Query Match 100.0%; Score 1043; DB 22; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.1e-101;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKIYKVPWMMAMGLVLSACSTPIPPKNNPOLAQTQTALSLDMGKLDQAQOQDAA 60
DB 1 MKIRVKKPMWMMGLVLSACSTPIPPKNNPOLAQTQTALSLDMGKLDQAQOQDAA 60
QY 61 LSADROPAPARTAKYVQASEDATHTKQRLFEKALINPKMOSYMDGFYVWMD 120
DB 61 LSADROPAPARTAKYVQASEDATHTKQRLFEKALINPKMOSYMDGFYVWMD 120
QY 121 LSGALIFDKPSRAIGYEGVVAIENMAYIYHGYEAKSPKDYNNAKSALERALISG 180
DB 121 LSGALIFDKPSRAIGYEGVVAIENMAYIYHGYEAKSPKDYNNAKSALERALISG 180
QY 181 TQHEIKKSYDKLSDYKLSDYK 204
DB 181 TQHEIKKSYDKLSDYKLSDYK 204

RESULT 2

AAW82501 ID AAW82501 standard; Protein; 1151 AA.

XX AAW82501;

DT 01-FEB-1999 (first entry)

XX C. elegans OGT protein.

XX OGT; O-linked GlcNAc transferase; uridine; transferase; tumour;
KM dihydrophospho-N-acetylglucosamine; polypeptide beta-N-acetylglucosaminyl;
KM predisposition; type II diabetes; hyperglycaemia; Alzheimer's disease;
KM metastasis; diagnosis; inhibitor; treatment; diabetes mellitus.

XX Caenorhabditis elegans.

OS Key Location/Qualifiers

FT MISC-difference 126

XX label= unknown
XX WO9844123-A2.
PN

XX 08-OCT-1998.

XX 27-MAR-1998; 98WO-US06101.

XX 31-MAR-1997; 97US-0042270.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Hanover JA, Lubas W;

XX WPI; 1998-557118/47.

DR N-PSDB; AAV69302.

PT Protein exhibiting O-linked GlcNAc transferase activity, OGT -
PT useful, e.g. to assess predisposition to type II diabetes or
PT Alzheimer's or metastatic potential of tumours, and to identify
PT inhibitors

PS Claim 14; Page 39-42; 56pp; English.

XX This sequence represents a Caenorhabditis elegans OGT, O-linked GlcNAc
CC transferase protein (also known as uridine dihydrophospho-N-acetylglucosamine:
CC polypeptide beta-N-acetylglucosaminyl transferase). This protein is
CC useful to assess predisposition toward type II diabetes in patients
CC suspected of having hyperglycaemia that could evolve into this disease,
CC by assaying OGT activity in red blood cells. It can also be used to
CC assess predisposition toward Alzheimer's disease, to assess the
CC metastatic potential of tumours and to diagnose a tumour with metastatic
CC potential. OGT can also be used to identify OGT inhibitors, especially in
CC high-throughput assays, useful, e.g. in the treatment of diabetes
CC mellitus, tumour-derived diseases and Alzheimer's disease.

XX Sequence 1151 AA;

Query Match 11.9%; Score 124.5; DB 19; Length 1151;
Best Local Similarity 27.3%; Pred. No. 0.00095;
Matches 44; Conservative 28; Mismatches 64; Indels 25; Gaps 6;

QY 30 NPQLAIRQIAISLDMGKLDQAQOQDAAALSADROPAPARTAKYVQASD---ATH 86
DB 257 NPDLVYCRSDGNLTKAMGRLERAKVCYLKALETOPQAVAMSLCCVFNQCEIWLAIH 316
QY 87 QTRQRLFEKALINPKMOSYMDYGFYVWMDLSGALIFDKPSRAIGYEGVVAIE- 145
DB 317 H-----FEKAVTLDPNPDAYINLGNVLEKARIPDRAVSAY---LRALNLSGNHVAVHG 367
QY 146 NMAYIYHQ-----YEA--KSPTKDYNNAKSALE 174
DB 368 NLACVYVEOGLDLAIDTYKKAIDLPHPFDACYNLANALK 408

RESULT 3

AAV75451 ID AAV75451 standard; Protein; 253 AA.

XX AAV75451;

DT 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 708 protein sequence SEQ ID NO:2376.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy.

XX Neisseria meningitidis.

OS WO9957280-A2.

PN 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.
PF

01-MAY-1998; 98US-0083758.
31-JUL-1998; 98US-0094869.
02-SEP-1998; 98US-0098994.
02-SEP-1998; 98US-0099062.
09-OCT-1998; 98US-0103749.
09-OCT-1998; 98US-0103794.
09-OCT-1998; 98US-0103766.
25-FEB-1999; 99US-0121528.
(CHIR) CHIRON CORP.
(GENO-) INST GENOMIC RES.
Frazer C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
Peterson J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
Tettelin H, Venter JC;
MPI, 2000-062150/05.
N-P8DB; AA254213.
 novel Neisserial polypeptides predicted to be useful antigens for
vaccines and diagnostics -
Claim 2; Page 1140-1141, 1453pp; English.
AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
PCR primers used in the exemplification of the present invention. The
polypeptides, the polynucleotides, antibodies and compositions of
the invention can be used as vaccines, as diagnostic reagents, and as
immunogenic compositions. The polypeptides can be used in the
manufacture of medicaments for treating or preventing infection due to
Neisserial bacteria (e.g. meningitis and septicemia), to detect the
presence of *Neisseria* bacteria, or to raise antibodies. They may also
be used to screen for agonists or antagonists, which may themselves
have use as antibacterial agents. The polynucleotides of the invention
may also be used in gene therapy protocols.

Query Match	11.3%	Score 117.5;	DB 21;	Length 253;
Best Local Similarity	23.6%	Pred. No. 0.00059;		
Matches	43;	Conservative	42;	Mismatches 78; Indels 19; Gaps 5

QY	4	RVKRPMVNMAGLIVLASCOST--PIPPKNNPQLAQRTQTQIAISLLDMGKLDAQOGLDAL	61
	:	: : : : : :	:
	:	: : : : : :	:
DB	8	RISLLLVIALTG---ACSTSYSPRAEKAANQSNTKTOLAMEVMNGODRYQAATSIDAL	63
	:	: : : : : :	:
QY	62	SADROFAPAYRTLAKYVOASEDATHQTKAQRLEFEKAIEINPKDMOSYMDGYFLY-OMGD	120
	:	: : : : : :	:
	:	: : : : : :	:
DB	64	KSDPNELAMLVBAEIYOYLK---VNDRKQSEFRRLAISIKPDASAINNNGWFLGGRILNR	120
	:	: : : : : :	:
QY	121	LSGALIITYDKSRRAIGYEGRVVAIENMAYIIYHQVEAAKSPTKDDYNNAKSLERALLSG	180
	:	: : : : : :	:
	:	: : : : : :	:
DB	121	PAESMAVDFKALADLPETPYTIANLN-----KGICSAKQGFGLAEAALKRSLLAAQ	171
	:	: : : : : :	:
	:	: : : : : :	:
QY	181	TQ 182	
DB	172	PQ 173	

RESULT 4
ABBS9330

ABB59330 ID	ABB59330 standard; Protein; 1059 AA.
XX	ABB59330;
XX	AC
XX	DT
XX	26-WAR-2002 (first entry)
DE	Drosophila melanogaster polypeptide SEQ ID NO 4782.
XX	
KM	Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX

XX WO200171042-A2.

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

XX

XX 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE) PE CORP NY.

XX

XX Venter JC, Adams M, Li FWD, Myers EW,

XX

XX WPI; 2001-656860/75.

XX

XX N-PSDB; ABL03433.

XX

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX

XX interactions -

XX

XX Disclosure; SEQ ID NO 4782; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

XX

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX

XX useful in developmental biology and in elucidating cell signalling and

XX

XX cell-cell interactions in higher eukaryotes for the development of

XX

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX

XX discloses genomic DNA sequences (ABL16176-ABL0511), expressed DNA

XX

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX

XX (AAB5773-AB572072).

XX

XX The sequence data for this patent did not form part of the printed

XX

XX specification, but was obtained in electronic format directly from WIPO

XX

XX at ftp.wipo.int/pub/published_pct_sequences.

XX

XX

XX Sequence 1059 AA;

Query Match	11.1%;	Score 116;	DB 22;	Length 1059;		
Best Local Similarity	24.8%;	Pred. No. 0.0067;				
Matches	50;	Conservative 37;	Mismatches 83;	Indels 32;		
				Gaps 7;		
Oy	18	SACSTDPPEKKNPQLAQIRQTALISLDMGKLDQAKQOLDALISADROFAPAYRTAKV	77			
Db	170	SAVAVAYTTLAQYNNPDLCYVRSDLGNLTKALGRLEEAACVYLKALETCPGFVAMSNIGCV	229			
Oy	78	YQASED---ATHQTKAQLRFEKAIENLPKXQMSYMDYGFVLYQMGDSLGAIIYFDKPSRA	134			
Db	230	FNAGGSEIWLAIHH-----PEKAVTLDPNFIADATINIGNNIKEKRIEDRAAAY---LRA	280			
Oy	135	IGYE-GRVVAIENNAVITYHQ-----YEA--KSPTKDDYNNAKSLERALISGT	181			
Db	281	LNLSPPNNAVHGNLACVYEGQLIDLAIDYFRRAIIEIQPNFPDPAVCNLIANLAK-----	333			
Oy	182	QHDEIKKSYDKLSDYKLSDY	203			
Db	334	EKGQVKEADECVNTALRLCSNH	355			
RESULT 5						
ABB67407						
ID	ABB67407	standard; Protein; 1059 AA.				
XX	ABB67407;					
AC						
XX	26-MAR-2002	(first entry)				
DT						
XX	Drosophila melanogaster	polypeptide SEQ ID NO 29013.				
DE						
XX	Drosophila; developmental biology; cell signalling; insecticide;					
KW	pharmaceutical.					

```

XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW,
XX DR WPI, 2001-656860/75.
XX DR N-PSDB; ABL11511.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 29013; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
XX CC sequences (AB101840-AB116175) and the encoded proteins
XX CC (AB101840-AB116175).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1059 AA;
XX
Query Match 11.1%; Score 116; DB 22; Length 1059;
Best Local Similarity 24.8%; Pred. No. 0.0067;
Matches 50; Conservative 37; Mismatches 83; Indels 32; Gaps 7;
QY 18 SACQSTPIPPKNNPQIAQIRTOIAISLDMGKIDQAKQQLDAISADROFAPAYRTLAKY 77
DB 170 SAVQAVITLQVNPDLVCVRSDGNLTKALGRLEBAKACYLKAIETCPGFAVAMSNIGCV 229
QY 78 YQASED---ATHQTKAQLFEKAIELNPKDMOSYMDYGYLVQMGDLGALIFYDKPSRA 134
DB 230 FNAQGEIWLAIHH-----FEKAVTLDPNPLDAVINLGNVLKEARIFDRAVAAY---LRA 280
QY 135 IGVE-GRVVAIENMAVIYHO-----YEAA---KSPTKDYNNAKSALERALLISGT 181
DB 281 LNI SPNNNAVHGNLACVYEEQGLIDLAIDTYRRAIELQNPFPAYCNLANALK----- 333
QY 182 QHDEIKKSYDKLSDYKLSIDY 203
DB 334 EKGQVKEAEDCYNTALRLCSNH 355

```

RESULT 6
ID ABB67408 standard; Protein; 1059 AA.
XX ABB67408;
XX AC
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 29016.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.

```

XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW,
XX DR WPI, 2001-656860/75.
XX DR N-PSDB; ABL11511.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 29016; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
XX CC sequences (AB101840-AB116175) and the encoded proteins
XX CC (AB101840-AB116175).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1059 AA;
XX
Query Match 11.1%; Score 116; DB 22; Length 1059;
Best Local Similarity 24.8%; Pred. No. 0.0067;
Matches 50; Conservative 37; Mismatches 83; Indels 32; Gaps 7;
QY 18 SACQSTPIPPKNNPQIAQIRTOIAISLDMGKIDQAKQQLDAISADROFAPAYRTLAKY 77
DB 170 SAVQAVITLQVNPDLVCVRSDGNLTKALGRLEBAKACYLKAIETCPGFAVAMSNIGCV 229
QY 78 YQASED---ATHQTKAQLFEKAIELNPKDMOSYMDYGYLVQMGDLGALIFYDKPSRA 134
DB 230 FNAQGEIWLAIHH-----FEKAVTLDPNPLDAVINLGNVLKEARIFDRAVAAY---LRA 280
QY 135 IGVE-GRVVAIENMAVIYHO-----YEAA---KSPTKDYNNAKSALERALLISGT 181
DB 281 LNI SPNNNAVHGNLACVYEEQGLIDLAIDTYRRAIELQNPFPAYCNLANALK----- 333
QY 182 QHDEIKKSYDKLSDYKLSIDY 203
DB 334 EKGQVKEAEDCYNTALRLCSNH 355

```

RESULT 7
ID AAY75450 standard; Protein; 253 AA.
XX AAY75450;
XX AC
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria gonorrhoeae ORF 708 protein sequence SEQ ID NO:2374.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX antibiotic; gene therapy.

OS Neisseria gonorrhoeae.
 XX
 PN MO957280-AZ.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098894.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tectelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR N-PSDB; AA254212.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 1140; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA254941
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 SQ Sequence 253 AA;
 Query Match 10.7%; Score 111.5; DB 21; Length 253;
 Best Local Similarity 23.1%; Pred. No. 0.0025;
 Matches 42; Conservative 43; Mismatches 78; Indels 19; Gaps 5;
 QY 4 RYKMPVWAMGLVLSACOST--PIPPKNNPQLAQRTOAISLLDNGKLDQAKQOLDAL 61
 DB 8 RISLLLVLLALG---ACSTSYRPSRAEKANQVSNITQLAMEYMRGCDYQATASIEDAL 63
 QY 62 SADRQAPAVRYTLAKYVQASSEDATHQTKAQLFEKAIELNPKMQSYMDYGFYLV-QMGD 120
 DB 64 KSNPKMELMALVRAEYIYQLK---VNDKQDESTRQALSTPDSAEINNNYGMFLCGLNR 120
 QY 121 LSGALTYFDKPSRAIGYGRVVAIENNAVYYHYQVEAKSPTRDQDYNNAKSALERALISQ 180
 DB 121 PASEMAVYFDKALADPYYPTFYIANLN-----KGICAKGQGFGLAEAYIKRSLAAQ 171
 QY 181 TQ 182
 DB 172 PQ 173

RESULT 8
 ABP38764
 ID ABP38764 standard; Protein; 230 AA.
 XX

AC ABP38764;
 XX
 DT 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3609.
 XX
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN US6380370-B1.
 XX
 PD 30-APR-2002.
 XX
 PF 13-AUG-1998; 98US-0134001.
 XX
 PR 14-AUG-1997; 97US-055779P.
 PR 08-NOV-1997; 97US-064964P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Doucette-Stamm LA, Bush D;
 PI
 XX
 DR WPI; 2002-381255/41.
 DR N-PSDB; ABN91309.
 XX
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX
 PS Disclosure; SEQ ID 3609; 267pp; English.
 XX
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 CC
 SQ Sequence 230 AA;
 Query Match 10.2%; Score 106.5; DB 23; Length 230;
 Best Local Similarity 27.5%; Pred. No. 0.0075;
 Matches 36; Conservative 25; Mismatches 53; Indels 17; Gaps 5;
 QY 75 ARVYQASEDATHQTKAQLFEKAIELNPKMQSYMDYGFYLVQMGDLGALTYFDKPSRA 134
 DB 13 SPIYQNTKQKIDTALKDIFEN-IEENPAIVENYINAGIVLSDVGIEKAEKPPQALTI 71
 QY 135 IYEGGRVVAIENNAVYYHYQVEAKSPTRDQDYNNAKSALERALISGTHDEIKKSYDKLL 194
 DB 72 EENQAV--YNNLANIYNE-----ERNEPAIKLYQTL-----QVEVAKKDCNWI 116
 QY 195 S-DYKLSIDYK 204
 DB 117 GMSFNQLGAFK 127

RESULT 9
 AAM93717
 ID AAM93717 standard; Protein; 408 AA.
 XX
 AC AAM93717;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 3663.
 XX

Query Match	Best Local Similarity	Score	DB 22;	Length 408;
Matches 50;	Conservative 29;	Mismatches 58;	Indels 73;	Gaps 8
25 IPPKNNPQLAQIRTOIAISILMDGKLDQAKQADALISA-----DROFAP-----	69			
162 IVDENLKDPAHAKSVTSCTLYNGIKLYHEQGHEALSYKEAIQKMPRFAPOSLYNM 221				
70 --AYRTLAKYQAS-----EDATH-----QTKARLFEKAIELNPK 103				
222 GSAVYRLSLKLPAAEHYMWESLRSKTDHIPALHTYGLKALALTGRKSEKFLPAIIEIDPT 261				
104 DMQSTYMDYFYLYQMDDLSGALIFYDPKPSAIGIEGRVVAIENMAIYYIHQYPAK--SP 161				
282 KGNCCVMHYQOFLIE-----EARLIEAAEVA-----KRAAEIDST 315				
162 TKDDYNNAKSALERALISGTHQHEIKKSYD 191				
316 EFDVVFNNAAHMLRQASL-----NEAEKTYD 341				

KW	Gibberellin signal transduction; spindly phenotype; SPY gene; rescue;
KW	SPY mutant gene, gibberellin overdose syndrome; paclobutrazol; spy-4 DNA
KW	modulation; plant development; plant height; fruit growth;
KW	flower development; leaf size.
XX	
OS	Arabidopsis thaliana.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..76
FT	/note= "N-terminus"
FT	77..430
FT	/note= "Tetracopeptide repeat region, which is
FT	important for SPY protein function"
FT	
FT	Region
FT	/note= "Tetratricopeptide repeat 1"
FT	111..151
FT	/note= "Tetratricopeptide repeat 2"
FT	152..185
FT	/note= "Tetratricopeptide repeat 3"
FT	186..219
FT	/note= "Tetratricopeptide repeat 4"
FT	220..260
FT	/note= "Tetratricopeptide repeat 5"
FT	261..294
FT	/note= "Tetratricopeptide repeat 6"
FT	295..328
FT	/note= "Tetratricopeptide repeat 7"
FT	329..362
FT	/note= "Tetratricopeptide repeat 8"
FT	363..396
FT	/note= "Tetratricopeptide repeat 9"
FT	397..430
FT	/note= "Tetratricopeptide repeat 10"
FT	431..914
XX	/note= "C-terminus"
XX	
PN	NO9743419-A2.
XX	
PD	20-NOV-1997.
XX	
PF	16-MAY-1997; 97WO-US06765.
XX	
PR	16-MAY-1996; 96US-0649046.
XX	
PA	(MINU) UNIV MINNESOTA.
XX	
PI	Jacobsen SE, Olszewski NE;
XX	
DR	WPI, 1996-008686/01.
DR	N-PSDB; AAV05171.
XX	
PT	New isolated spindly gene from plants - is involved in gibberellin
PT	signal transduction, used to develop products for altering plant
PT	development
XX	
PS	Disclosure; Fig 2; 54pp; English.
XX	
CC	The present sequence represents a protein that is involved in gibberellin
CC	signal transduction. Inactivation of the gene produces a spindly
CC	phenotype (hence the SPY gene). The spindly mutation is characterised by
CC	elongated petioles, yellow-green leaves, early flowering, long spindly
CC	bolts, partial male sterility and parthenocarpic fruit development. These
CC	phenotypes are also observed in wild type plants exhibiting a gibberellin
CC	overdose syndrome due to external application of gibberellin. A spy
CC	phenotypic mutant was isolated from a library of Agrobacterium-mediated
CC	seed transformation lines of Arabidopsis having T-DNA insertions
CC	positioned throughout the genome. Seeds from the library were selected
CC	for their ability to germinate in the presence of paclobutrazol.
CC	Restriction mapping was used to identify spy mutant (spy-4) DNA from
CC	resistant seeds. This DNA was used as a probe to obtain the spy cDNA.
CC	Introduction of the SPY gene into plants rescues the spindly phenotype.
CC	The SPY DNA, vectors and proteins can be used to modulate plant
CC	development including plant height, fruit growth, flower development and
CC	development

CC leaf size.
XX
SQ Sequence 981 AA;

Query Match 10.2%; Score 106.5; DB 19; Length 981;
Best Local Similarity 23.7%; Pred. No. 0.06; Mismatches 76; Indels 43; Gaps 7;
Matches 45; Conservative 26; Mismatches 76; Indels 43; Gaps 7;

QY 33 LAQIRFOIAISLDMGKLDQAKQOOLDAASDRQFAPAYRTAKVYQASEDAHTQTKAQR 92
DB 218 LAVALTDLGTSLSKAGTQEGCIQYKYEALKIDPHYAPAYNLGVY---SEMQYDNALS- 274
QY 93 LFEKALIELNPKDMQSYMDYGFYLVQMGDSLALYEDK-----PSRAIGYGRVAI--- 144
DB 275 CYEKALERPMVAEAVCNMGVIYKNGDLEMAITCYERCLAVSPNEIANKNNMALTLTL 334
QY 145 -----ENMAY-----ITYHQYBAKSPFKDDYNNAKSLERALLISGTHDEIKS 189
DB 335 GTRVKLEGDVLTQGVAAVYKKALYYNMHYADAM-----YN-----LGVAHGEWLK- 377
QY 190 YDKLSDYKL 199
DB 378 FDMAIVPYEL 387

RESULT 11

AAV32160
ID AAV32160 standard; Protein; 264 AA.

AC AAV32160;

DT 01-FEB-2000 (first entry)

XX Soybean CDC-16 protein fragment.

XX CDC-16; soybean; cell cycle regulatory protein;

KW transcription factor; herbicide.

OS Glycine max.

XX MO9953075-A2.

PN 21-OCT-1999.

XX 08-APR-1999; 99MO-US07638.

PF 09-APR-1998; 98US-0081132.

PR (DUPO) DU PONT DE NEMOURS & CO E. I.

PA Klein TM, Morakinyo LO, Odell JT, Sakai H;

PI WPI, 1999-633830/54.

XX N-PSDB; AA234576.

XX Plant-derived cell cycle regulatory proteins -

PS Claim 5; Page 34-35; 44pp; English.

XX This sequence represents 43% of the middle region of soybean
CC cell cycle regulatory protein CDC-16, as deduced from an isolated
CC cDNA clone (see AA234576). The invention relates to nucleic acid
CC fragments (see AA234575-83) encoding plant CDC-16, DP-1, DP-2 and
CC BPF cell cycle regulatory proteins (see AA23259-67). It also
CC relates to the construction of a chimeric gene encoding all or a
CC portion of the cell cycle regulatory protein, in sense or antisense
CC orientation, where expression of the chimeric gene results in
CC production of altered levels of the cell cycle regulatory protein in
CC a transformed host cell. The nucleic acid and proteins may be
CC used to facilitate studies of cell cycle regulation in plants,
CC provide genetic tools to enhance cell growth in tissue culture,
CC increase gene transfer efficiency and provide more stable
CC transformations. The proteins may also provide targets to

CC facilitate design and/or identification of cell cycle regulatory
CC proteins that may be useful as herbicides.

XX
SQ Sequence 264 AA;

Query Match 10.2%; Score 106; DB 20; Length 264;
Best Local Similarity 23.9%; Pred. No. 0.01; Mismatches 89; Indels 56; Gaps 9;
Matches 54; Conservative 27; Mismatches 89; Indels 56; Gaps 9;

QY 15 LVLSAQSTPIPPKNNPQLAQIRTOAISLDMGKLDQAKQOOLDAASDRQFAP----- 69
DB 34 LYIMSCNLV---KDPQMALSMFPAVGCYYCICKYDQSRRYFSKXTSLDGTFFPAMIGY 89
QY 70 -----AYRTLAKVYQASEDA-----THQTK-AQRLFEKALIELNPK 103
DB 90 GNAYVAQEBGDQAMSVIRTRARLFPQCHLATLYIGMECHRTHSYKLABQFFQAKSICSS 149
QY 104 DMQSYMDYGFYLVQMGDSLALYEDKPSRAIG-----YEGRVVAIENMAYIYHQYEA 157
DB 150 DPLVYVELGVAVYHMEYKKAVVMFEKTLALVPTTLSIWEESTV---NLA---HAYRK 202
QY 158 AKSPTKDDYNNAKSALERALLISGTHDEIKSYDKLSDYKLSDY 203
DB 203 LKQ-----YRAISYKALALSTRS---VSTYAGLAYTYHLODDP 240

RESULT 12

AAV75452
ID AAV75452 standard; Protein; 253 AA.

AC AAV75452;

DT 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 708 protein sequence SEQ ID NO:2378.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

XX antibacterial; gene therapy.

OS Neisseria meningitidis.

XX MO9957280-A2.

PN 11-NOV-1999.

XX 30-APR-1999; 99MO-US09346.

PF 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

PI Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizsa M, Rappold R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX WPI, 2000-062150/05.

XX N-PSDB; AA254214.

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

PS Claim 2; Page 1141; 1453pp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941

CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria meningitidis* and *Neisseria gonorrhoeae*, to detect the
 CC presence of *Neisseria meningitidis*, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

SQ Sequence 253 AA;

Query Match 9.9%; Score 103.5; DB 21; Length 253;
 Best Local Similarity 22.0%; Pred. No. 0.018;

Matches 40; Conservative 42; Mismatches 81; Indels 19; Gaps 5;

QY 4 RVKMPVWANGVLISACQST--PIPPKNNPOLAQRITQIAISLIDMGKLDQAKQDLAL 61
 8 RISLILVLAIG---ACSTSYRPSRAKAKANQVSNIKTOLAMEYWRGODYRQXTASTEDAL 63

DB 62 SAARQFAPAYRTIAKAVYQASEDATHQTKAQLFEKALELNPQKQSYMDYGFYLV-OMGD 120
 64 KSDPKELAMLVRAEITQYK--VNDKAQSEFQXISIRFDSAEINNNYXWFLCGRINR 120

QY 121 LSGALIFDKPSRAIGYGRVALENNMAYIYHQYEAKSPKDYNNAKSALERALISG 180

DB 121 PAESMAYFDKALADPTYPXPYIANLN-----KGISCAKQGFGLAEAYLKRSLAAQ 171

QY 181 TQ 182
 DB 172 PQ 173

RESULT 13

ABBA9001
 ID ABB49001 standard; Protein: 212 AA.

AC ABB49001;

DT 05-FEB-2002 (first entry)

DE *Listeria monocytogenes* protein #1705.

KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KM vitamin B12; bacterial infection; disease.

OS *Listeria monocytogenes*.

PN W0200177335-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-FR01118.

PR 11-APR-2000; 2000FR-0004629.

PA (INSP) INST PASTEUR.

PI Buchrieser C, Frangoul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussauget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tietz-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baguerio F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablos B, Wehlend J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;

PA WPI; 2002-010914/01.

Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment

PT and prevention of *Listeria* and related bacterial infections, and
 PT related polypeptides -
 XX Claim 6; SEQ ID No 1706; 192pp; French.

XX The present invention relates to the genome sequence of *Listeria*
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in *L.*
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of *L. monocytogenes* and related organisms,
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by *L.*
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 212 AA;

Query Match 9.7%; Score 101.5; DB 23; Length 212;
 Best Local Similarity 23.8%; Pred. No. 0.022;

Matches 41; Conservative 24; Mismatches 66; Indels 41; Gaps 4;

QY 44 LIDMGKLDQAKQDLAALSADROFAPAYRTIAKAVY-----QASEDATHQ----- 87

DB 34 LISMDFERAELEFKRALILDITVPAAYSLNLVYELRYOEADSFQATKQGMENGD 93

QY 88 -----TKAQLFEKALELNPQKQSYMDYGFYLVOMGDSGLIYFDKRS 132

DB 94 LPEMLGMSFVQMEELTAMPYLLRSVEINPEGEALFYGIYVARSQ-----FIED--- 144

QY 133 RAIGYGRVALENNMAYIYHQYEAKSPKDYNNAKSALERALISGTQHD 184

DB 145 -ALNMLERVLVKEPDDPALYNIGAAVYLMQGDIVAKVYFERALITGASHE 195

RESULT 14

AAU32898
 ID AAU32898 standard; Protein: 230 AA.

AC AAU32898;

DT 18-DEC-2001 (first entry)

DE Novel human secreted protein #3389.

KM Human; vaccination; gene therapy; nutritional supplement;
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.

PN W0200179449-A2.

PD 25-OCT-2001.

PF 16-APR-2001; 2001WO-US08656.

PR 18-APR-2000; 2000US-0552929.

PA 26-JAN-2001; 2001US-0770160.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

PA WPI; 2001-611725/70.

XX Venter JC, Adams M, Li FWD, Myers EW,
XX WPI, 2001-656860/75.
DR N-PSDB; ABL12017.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila*, and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 30534; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB101840-AB16175).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 486 AA;
Query Match 9.2%; Score 96; DB 22; Length 486;
Best Local Similarity 26.4%; Pred. No. 0.28;
Matches 39; Conservative 28; Mismatches 63; Indels 18; Gaps 5;
OY 64 DROFPAVRLTAKYQASBETHQKARLEKALINPKMDGYFLVQMGDLSG 123
DB 61 NRHLNPEFLYVQGLIDREBGNH-LEALRHUQSAELPNRITETKEIGRLTYMGKRSQ 119
OY 124 ALIYFDKSRRAIGEGVVAIENMAYIYHQ--YEAKSPKTD-----YNNAKSALERA 176
DB 120 ALGVREKRA-----EGRSSRDHEIYHVLGELLYAAATTOGQKDVASQODEARTYELA 173
OY 177 LISGTQHDEIKKSYDKLISDYKLSDYK 204
DB 174 VQSGRK-----LESYVRLAELVYKDKQYQ 197
RESULT 17
AAU99188 standard; Protein; 585 AA.
XX
AC AAU99188;
XX
DT 24-SEP-2002 (first entry)
XX
DE Target molecule human PAT1.
XX
KW Retinoid acid pathway; RA; retinoid; lung cancer;
KW Kaposi's sarcoma; breast cancer; pancreatic cancer; neuroblastoma;
KW renal cancer; ovarian cancer; dermatitis; hyperkeratosis; eczema;
KW Darier's disease; Reiter's disease; psoriasis; gene therapy; R2;
KW acute promyelocytic leukemia; APL; Perturbagen; R3; F802; F820;
KW yeast two-hybrid assay; PAT1; kinesin light chain-related protein;
KW human.
XX
OS Homo sapiens.
XX
FN WO200240719-A2.
XX
PD 23-MAY-2002.
XX
PF 17-NOV-2001; 2001WO-US44039.
XX
PR 17-NOV-2000; 2000US-249468P.
XX
PA (DELT-) DELTAGEN PROTEOMICS INC.
XX

PI Kamb CA, Richards BT, Karpilow J;
XX WPI, 2002-519386/55.
DR N-PSDB; ABR87389.
XX
PT Polypeptide with retinoid acid pathway activity, especially of
PT perturbagens R3, F802 and F820 for identifying a cellular target which
PT interacts with the polypeptide and for therapeutic purposes -
XX
PS Claim 93; Fig 16d; 131pp; English.
XX
CC The invention relates to an isolated polypeptide (I) with retinoid acid
CC (RA) pathway activity, comprising a polypeptide sequence of Perturbagen
CC (phenotypic probe) R3, perturbagen F802, perturbagen F820, their
CC biologically active modifications, or biologically active fragments.
CC Also included are the polynucleotides encoding the perturbagens, a
CC vector comprising the polynucleotide, preparing an RA pathway related
CC polypeptide, a composition comprising the polypeptide, an antibody to the
CC polypeptide, screening (M) putative RA-related therapeutics, by exposing
CC a polypeptide/target interaction pair identified by a method utilising
CC the RA pathway polypeptide, to a number of agents, and recovering a
CC subpopulation of disrupting agents which competitively displace the
CC polypeptide from the target, where the disrupting agents are putative RA-
CC related therapeutics. Also include are an isolated RA pathway polypeptide
CC comprising PAT1 (a kinesin light chain-related protein) polypeptide
CC and its encoding polynucleotide, a gene therapy vector comprising the RA
CC pathway protein polypeptide or encoding or PAT1 and a host cell
CC comprising the gene therapy vector. The RA pathway polypeptide is useful
CC for identifying a cellular target that interacts with RA pathway-related
CC molecules and identifying a polypeptide/target interaction pair, by
CC detecting reporter expression, where the reporter expression is
CC operatively linked to the formation of the interaction pair. (M) is a
CC yeast two-hybrid assay. The polypeptide is also useful for treating
CC an RA pathway-related condition e.g. lung cancer, Kaposi's sarcoma,
CC breast cancer, pancreatic cancer, neuroblastoma, renal cancer,
CC ovarian cancer, dermatitis, hyperkeratosis, eczema, Darier's disease,
CC Reiter's disease, psoriasis, acute promyelocytic leukemia (APL).
CC The present sequence is a target molecule isolated by the yeast two
CC hybrid system, PAT1.
XX
SQ Sequence 585 AA;
Query Match 9.1%; Score 95; DB 23; Length 585;
Best Local Similarity 25.7%; Pred. No. 0.47;
Matches 45; Conservative 24; Mismatches 56; Indels 50; Gaps 8;
OY 47 MGLDQAKQ-----LDALSDROFAPAYR-----TLAKYQASBETHQRT 88
DB 193 MDKLKSKHQQAANKALYGLCALIFAKSHYDEAKYKCIAMKEITGLPVKVVVDLRLQA 252
OY 89 -----KAORLEKAIEL-----NPKDMOSTVMDGYFLVQMGD--SGALLY 127
DB 253 SKACVVRBEFKKAEQILKHAIVLARDHFGSKHKPYDITLDYGFYLLNVDNTIQSAIVQ 312
OY 128 FDKPSRAIGEGR-----VAIENMAY-IYHVOEAKSPKTDYNNAKSALERA 177
DB 313 AALDIRQSVGKNIHVATNHEDLAVSSYVHQSCK-----FDNALFRAERAI 361
RESULT 18
AAG24834 standard; Protein; 333 AA.
XX
ID AAG24834;
XX
AC AAG24834;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28659.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

XX Arabidopsis thaliana.
OS
XX BP1033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130510.
PR 22-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138867.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144326.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149932.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

```
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158923.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160982.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 8.9%; Score 93; DB 21; Length 333;
Best Local Similarity 25.9%; Pred. No. 0.34; Mismatches 87; Indels 34; Gaps 7;
Matches 51; Conservative 25; Mismatches 87; Indels 34; Gaps 7;

```
QY 12 AMGVLSACGCTPIPPKNNPOLAOTFOIAISLIDMG---KLDQAKQDLDAISADROFA 68
DB 74 AAALFLVNLQKPSPAIAAPVAA---TPIMESLKQSGNNGVNFEEERSLREYLASHDDV 130
QY 69 PAYRTLAKVVOASEDATHQTKARLFKAEIENPKDMQSYMDYGFYVQMGDLGSLIYF 128
DB 131 EALRSLMEVRIKSKKL---LEAFLIDRLIELEFEKEWEPULKANIFSYSGDLESATGF 187
QY 129 DK-----PSRAIGYEGRAVAIENNAVITYHQYEAKSPTPDDYNNAKSALERALLISGTQH 183
DB 188 EEIIVKDPLEAVEAHGLV-----MAY-----SDSGDDINAAVEQRIEAMVCK-- 230
QY 184 DEIKKSYDKLSDDYKL 200
DB 231 ---KEKNRKDIDRFKLL 244
```

RESULT 19
AAC43345
ID AAC43345 standard; Protein: 333 AA.
XX AAC43345;
AC
XX
DT 18-OCT-2000 (first entry)
XX

```
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54166.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 16-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 17-JUN-1999; 99US-0139453.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
```


XX AC AAG43344;
XX AC 18-OCT-2000 (first entry)
XX DT
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 54165.
XX DE
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX OS EPI033405-A2.
XX PN
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PF 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136382.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139452.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142054.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

```

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match Best Local Similarity 8.9%; Score 93; DB 21; Length 344;

Matches 51; Conservative 25; Mismatches 87; Indels 34; Gaps 7;

```

QY 12 ANGVLVSACGCTTTPPKNNPQALQITQIALISLDNG--KLDQAKQQLDPAISADRORA 68
DB 85 AALFLVNIQLTKPSPIAIPVAA---TPLMESLKOSNGNVSPFEERSLFEVYASHPPDV 141
QY 69 PAVRTLAKYQASBEDATHQKARLFEKALINPKXMOQSYNDYGFVLVQMGDISGLIYF 128
DB 142 EAKRSLMEVRIKSRKI---LEALELIDRLIELEPEEKEMWMLKANIFSYSGDLESAKTGF 198
QY 129 DK-----PSRAIYEGRVVAIENMAYIYHQVEAKSPTRQDYNNAKSALERALLISGTOH 183
DB 199 EELIVNDPLRVEVYHGV-----MAY-----SDSGDDLANAVEKRIEEMVAVCK-- 241
QY 184 DEIKKSYDKLLSDPKYL 200
DB 242 ---KERNRKDLRDFKLL 255

```

```

RESULT 21
ID AAG24833
XX AAG24833 Standard; Protein; 345 AA.
AC AAG24833;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28658.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
EF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137702.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139493.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.

```


CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.

XX Sequence 860 AA;

Query Match 8.9%; Score 92.5; DB 22; Length 860;

Best Local Similarity 25.5%; Pred. No. 1.5; Mismatches 69; Indels 7; Gaps 4;

Matches 35; Conservative 26; Mismatches 69; Indels 7; Gaps 4;

DB 28 KNNPOLAQRTOAISLDMGKLDQAKQQLDPAALSDRQAPAPAYRTLAQVQASEDAHQ 87

DB 280 KNGP-IAMLYK--GLTFPHRGLLKEALIESFKELKQKVDIFDAYKSGQAVRELFEEA 336

DB 88 TKQRLFEKALINPKMOSYMDYGFILVQMGDISGALITYDKPSRAIGYGRVALIEM 147

DB 337 TES---FOKALLINQNHVQTQLRGMMLYHSGLSQELKMKRCLQLEPNEVCQYMKGL 393

DB 148 AYIYHO-YEAKKSPTK 163

DB 394 SHVAMGQFYEGIKAKTK 410

RESULT 24

ABBS9359 ID ABB59359 standard; Protein; 490 AA.

XX ABB59359;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 4869.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-6556860/75.

XX N-PSDB; ABL03462.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure; SEQ ID NO 4869; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA

XX (ABBS7737-ABBS72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 490 AA;

Query Match 8.8%; Score 91.5; DB 22; Length 490;

Best Local Similarity 25.0%; Pred. No. 0.84; Mismatches 43; Indels 3; Gaps 1;

Matches 21; Conservative 17; Mismatches 43; Indels 3; Gaps 1;

DB 28 KNNPOLAQRTOAISLDMGKLDQAKQQLDPAALSDRQAPAPAYRTLAQVQASEDAHQ 87

DB 338 KNPDPKLYSNNAACYTTLAFLDGLKDCDTCIKDEKFIKGIKIKLQMQ---00 394

DB 88 TKQRLFEKALINPKMOSYMDY 111

DB 395 SKAQAPYQALIELDPNNAEAIKGY 418

RESULT 25

AAV79489 ID AAV79489 standard; Protein; 387 AA.

XX AAV79489;

XX 01-AUG-2000 (first entry)

XX Human testis-specific 22P4F11 protein.

XX 22P4F11; human; testis; prostate cancer; diagnosis; therapy;

XX marker; vaccine.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 11..16 /note="putative mitochondrial signal sequence"

XX WO200018925-A1.

XX 06-APR-2000.

XX 30-SEP-1999; 99WO-US3005.

XX 30-SEP-1998; 98US-0102572.

XX 28-JUL-1999; 99US-0146584.

XX (UROG-) UROGENESYS INC.

XX (AFAR/) AFAR D E.

XX (HUBE/) HUBERT R S.

XX (MITC/) MITCHELL S C.

XX Afar DE, Hubert RS, Mitchell SC;

XX WPI; 2000-303452/26.

XX N-PSDB; AA294893.

XX Novel testis-specific gene 22P4F11 which is expressed in human prostate

XX cancer and is useful as a diagnostic marker and/or therapeutic target

XX for prostate cancer -

XX Claim 1; Fig 1A; 54pp; English.

XX The present sequence is that of human 22P4F11, as deduced from

XX an isolated cDNA clone (see AA294893). Examination of 22P4F11 mRNA

XX expression in normal prostate, prostate tumour xenografts, and a

XX variety of normal tissues indicated that the expression of the

XX protein is testis specific in normal tissues. 22P4F11 is also

XX expressed in human prostate cancer xenograft tumours, in some cases

XX at high levels. The 22P4F11 transcript and/or protein may represent

XX a useful diagnostic marker and/or therapeutic target for prostate

XX cancer. The invention provides 22P4F11 polypeptides and

XX polynucleotides, methods for recombinant production, antibodies,

XX transgenic animals, assays for detecting the presence of 22P4F11

XX polynucleotide or mRNA in a sample, a method for detecting the

XX presence of a cancer expressing 22P4F11 protein, a method of

XX diagnosing the presence of cancer (especially prostate cancer),

XX a vaccine composition for treatment of cancer comprising a

CC 22P4F11 protein or immunogenic portion, and a method of inhibiting
CC the development of a cancer expressing 22P4F11 by administering
CC the vaccine (all claimed).

SO Sequence 387 AA;

Query Match 8.7%; Score 91; DB 21; Length 387;
Best Local Similarity 23.8%; Pred. No. 0.68;
Matches 59; Conservative 21; Mismatches 86; Indels 82; Gaps 11;

QY 30 NPLQAIQRIQIAISLDMGLDQAKOOLADALSADROFAPAYITLAKV-----YQASE 82
DB 144 NPAVYIARISFGVYLAQAGKFORAMNHFTIAIDTPKNVLAEGRAVVCLOMGNNPAAQ 203
QY 83 DAT-----HOTKARLFKFAILEPNKQSYMDYGFYLVQMG 119
DB 204 DINAANKISTTAFLNTRGVIEHFMGHKQAMKDYQALITLNRKYSLAYFNAGNIYFHHR 263
QY 120 DLSCALIYFDKP-----SRAI-----GYEGRVVAIE-----NM 147
DB 264 QFSQASDYFSEKALKPENEFYLMNRATITWILKKYEAKEDPANVIESCPMAAVYFNR 323
QY 148 AVIYY--HOYEAKSPKDDYNNAKS-----ALE--RALISGT--QHDEIKSYDKLS 195
DB 324 AHFYCYLKOYELA-----EEDLNKALSLKPNDAIVNFRKVRGKIGLIBAMADYNQALD 379
QY 196 DYKLSDPY 203
DB 380 ----LEDY 383

RESULT 26

AAG42251
ID AAG42251 standard; Protein: 465 AA.

AC AAG42251;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52669.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129465.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 99US-0132486.

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

```

PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 25-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145293.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147008.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147459.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148688.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

```

```

PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      8.7%; Score 90.5; DB 21; Length 465;
Best Local Similarity 22.1%; Pred. No. 1;
Matches 58; Conservative 34; Mismatches 74; Indels 97; Gaps 12;

QY 11 MAMGLVLR---SACSTPIPPKNNPQLAIRTOIAISLDMGKLDQAK----- 54
DB 1 MAYTLLILNFVPAQLLLIQ-----LVSAIDGQGVDAELFEFASOSIKV 47
QY 55 -----QQLDALISADROPAPVITLAKY-----YQASEDTHTQTKQRLFEKAIENLP 102
DB 48 RYSDALDDINAAIEADPALSEAVYFKRASVLRHFCREYESENS-----YQKTLERKS 98
QY 103 KMQSYMDGYFLVQWGLDLSGAL-----IYFDPK-SRAIGYGRVVAIENMA----- 148
DB 99 GDSNAKE-----LSQHQKSALEFASITYESKDIAKALEFQKVLVSPACSKAKLK 154
QY 149 ---YIYHQYEAKSPT---KDYNNAKSAL-----ERALLSGTQ 182
DB 155 VKLLWMSKDYSGAISETYGILKEDENNLEALLRGRAYVYLADHDIAQRHYQKGLRDE 214
QY 183 HDEIKKSY---DKLLSDYKLLSD 202
DB 215 HSELKRAYFGLKLLKTKYSABD 237

RESULT 27
AAG42250
ID AAG42250 standard; Protein; 473 AA.
AC AAG42250;
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52668.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127452.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.

```

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132464.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132853.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142350.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145213.
PR 27-JUL-1999; 99US-0145819.
PR 27-JUL-1999; 99US-0145819.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154179.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.7%; Score 90.5; DB 21; Length 473;

Best Local Similarity 22.1%; Pred. No. 1;

Matches 58; Conservative 34; Mismatches 74; Indels 97; Gaps 12;

QY 11 MAMGVL-----SACSTPIPPKNNPQIAIRFOAISLMDKLDQAK----- 54
DB 9 MAYTLILNFVPAQILLIIP-----LVSLDQGSVAALFERASQSIKVK 55
QY 55 -----QQLDALSDROPAPARYRLAKV-----YQASEDATHOTKQRLFEKAIELNP 102
DB 56 RYSDALDILNAIIEADPALSEAYFKASVLRHFCRYEDSENS-----YQKLEPKS 106
QY 103 KMOGYMVGFTLVQMGDLSCAL-----IYFKXP-SRAIGYBGRVAIENNA----- 148
DB 107 GDSNAKE-----LSQHQKSALETASTLYESKDIKAKALEFVDKVLVFSACKKAKLKK 162
QY 149 ---YIYHGYENAKSP-----KDYNNAKSAL-----ERALLISGTQ 162
DB 163 VKLIMVSKDYSALSETGYILKEDENNIEALLIRGRAYVYADHDIAQRHYQKGLRDP 222
QY 183 HDEIKSY---DKLISDYKLLISD 202
DB 223 HSELKKAAYFGKJLKKTKTSAED 245

RESULT 28

AAG42249 ID AAG42249 standard; Protein; 482 AA.

XX AAG42249;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52667.

KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301419.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132853.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142972.
PR 13-JUL-1999; 99US-0143547.
PR 14-JUL-1999; 99US-0143624.

```
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
```

```
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match

Best Local Similarity 22.1%; Pred. No. 1.1; DB 21; Length 482;
Matches 58; Conservative 34; Mismatches 74; Indels 97; Gaps 12;

```
QY 11 MAMGLV-----SACQSTPIPPKNNPOLAQIRTOIAISLDMGKLDQAK----- 54
DB 18 MATYLLINVFACQILLQF-----LVSAIDGQGVDAALFERSQSQIKVK 64
QY 55 -----QQLDALSLADROFAPAYRTLAKV-----YQASEDAHQTKAORLFKAILENP 102
DB 65 RYSDALDDINAAIEADPALSEAYFKRASVLRHRCRYEDSENS-----YQKYLEFKS 115
QY 103 KDMQSYMDYGFYLVQMGDLSGAL-----YFDPK-SRALCYEKRYVAIENMA----- 148
DB 116 GDSNAEKE---LSQDHQAKSALETASTLYESKDIAKALEFDKVVLPVSPACKKLK 171
QY 149 ---YIYHGYEAKSPKPT---KDYNNAKSAL-----EPALISGTQ 182
DB 172 VKLLMWSKQYSGAISTGYLIKEDENNLEALLRGAYYYIADHDIAQRHYQGLADPDE 231
QY 183 HDEIKKSY---DKLSDYKLLSD 202
DB 232 HSELKKAYFGLKKLLKTKKSAED 254
```

RESULT 29

ABP09759 ID ABP09759 standard; Protein; 165 AA.

XX AC ABP09759;

XX DT 24-JUN-2002 (first entry)

XX DE Human OREFX protein sequence SEQ ID NO:19500.

XX KW Human; open reading frame; OREFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumor; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;

KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
OS Homo sapiens.
XX MO200192523-A2.
PN 29-MAY-2001; 2001WO-US10836.
XX 06-DEC-2001.
PD 29-MAY-2001; 2001WO-US10836.
XX 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX (CURA-) CURAGEN CORP.
PA Shimketa RA, Leach MD;
XX WPI: 2002-106308/14.
DR N-PSDB; ABN25511.
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX Disclosure; SEQ ID 19500; 1037bp; English.
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic
CC transplantation, cardiovascular diseases, diabetes mellitus, cholesterol ester
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or peritoneal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIRO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 165 AA;
SQ

Query Match 8.6%; Score 90; DB 23; Length 165;
Best Local Similarity 23.8%; Pred. No. 0.25;
Matches 34; Conservative 36; Mismatches 53; Indels 20; Gaps 7;

QY 70 AYRTLAKYQASEDAHQTKQRLPEKAIENPKDMQSYMDYGYLVQMGDLGALTYFD 129
DB 1 AY-TLQGHSHSNDSDSAKT--CYRKALACDPQYHNNYVGGTSGAMKLGVEEALTYFE 57
QY 130 KPSRAIGYEGRVV-----AIENMAY-----IYHYEAKSPYTKDDYNNASALERALIS 179
DB 58 K-ARSINIEVNVVLLICCGSGLEKLGKSKALQYELACHLOPTS---SLSKYKMGQLLYS 113
QY 180 GTOHDEIKSYDKLSDYKLTSD 202
DB 114 MTRIVVALQTFEELV---KLVPD 133

RESULT 30
AAAG24835
ID AAAG24835 standard; Protein; 234 AA.
XX
XX AAAG24835;
AC
XX
XX 17-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28660.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EPI033405-A2.
FN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128334.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130691.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134378.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135333.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.

RESULT 31
AAG43346
ID AAG43346 standard; Protein: 234 AA.
XX AAG43346;
AC
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54167.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KM hydrolisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0131825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132046.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145193.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 8.6%; Score 90; DB 21; Length 234;

Best Local Similarity 25.8%; Pred. No. 0.42; Mismatches 63; Indels 28; Gaps 5;

```

Matches 40; Conservative 24;
DB 51 DQAKQDALSADROFAPARTLAKYQASEDAHTHTKQRLFEKAIELNPKMQSYMD 110
DB 14 EEERSRLEEVYASPPDVEALRSIMEVIRIKRKU--LEAIELDLRIELEPEEKEMPM 70
QY 111 YGFYLVQMDLSCALYIFDK-----PSRAIGEGRVVAIENMAYIYHYEAAKSPTKD 165
DB 71 KANIFSYSGLESKAKTGFEEELIVKDPLEVAHYGLV-----MAY-----SDSGDD 115
QY 166 YNNAKSALERALLSGTOHDEIKSYDKLSDYKLL 200
DB 116 LNAVEKRIEAMVRCK-----KEKNRKDLDFKLL 145

```

```

RESULT 32
AAR20568
ID AAR20568 standard; Protein; 900 AA.
AC AAR20568;
AC AAR20568;
DT 17-MAY-1992 (first entry)
DE Sequence of a partial polypeptide isolated from N. meningitidis
DE which is substantially homologous with a segment of a haemolysin
DE toxin in clone A4.85.
KM Vaccine; toxin; monoclonal antibody; probe; diagnosis; immunoassay;
KM meningitis; haemolysin.
OS Neisseria meningitidis strain FAM20.
FH Key Location/Qualifiers
FT Region 486..512
FT Region /label= haemolysin consensus sequence
FT Region 623..712
FT Region /label= see above
FT Region 823..900
FT Region /label= see above
PN MO9201460-A.
PD 06-FEB-1992.
PF 16-JUL-1991; 91WC-1005014.
PR 16-JUL-1990; 90US-0552649.
PR (UYNC-) UNIV NORTH CAROLINA.
PI Spalling PF, Thompson SA;
XX WPI, 1992-064698/08.
XX N-PSDB; AAQ21137.
XX DR
XX PT New antigenic polypeptide isolated from Neisseria meningitidis -
XX PT useful as vaccine against infections n.meningitidis diseases and
XX PT for detecting n.meningitidis in sample
XX PS Claim 6; Fig 2; 71pp; French.
XX CC The haemolysin-like toxin of the invention is isolated from
XX CC Neisseria meningitidis (N.m.) grown in vitro under iron-limiting
XX CC conditions (see AAR20568). It has 3 stretches of multiple repeats of
XX CC the haemolysin sequence and a total 21 complete consensus sequences
XX CC (see also AAR20569). MAb (A4.85) raised against AAR20568 was used to
XX CC screen a genomic library in lambda gt11 and DNA from the positive
XX CC clone identified was used to identify adjacent genomic fragments.
XX CC Together the fragments identified includes a 2721 bp open reading
XX CC frame although this did not include the start or the end of the gene
XX CC (see AAQ21137).
SQ Sequence 900 AA;

```

Query Match 8.6%; Score 90; DB 13; Length 900;

Best Local Similarity 23.2%; Pred. No. 2.9; Mismatches 69; Indels 90; Gaps 15;

```

Matches 58; Conservative 33;
DB 1 MKIRKWPMMVMAGLVLSACQSTPIPPKXNPQLAQIRTOIAISLDGK--LDQAKQ-- 56
DB 275 MRLSTDWOTANEGIAL-----TP-----SQVAQLKQNALVSLSDKAKAIDAARDRIA 323
QY 57 -LDAALSARQFAPARTLAKYQASEDA-----THQTKAQQLFEKAI---ELNPK 103
DB 324 VLDAYTGQSS-----TL--YMSSEDAINTVKNYDNTYDHLAKNIYQNLFFQTRLP-- 374

```

```

QY 104 DMOSYMDYGFYLVOMG-----DLSGALIYED-----KPSRAI----- 135
DB 375 -----YLNQISFKMENDFTTLPDSGLVOAFNHVETNPQKAFVDLAEMLAAGELR 424
QY 136 -GYEGRVAAIEMNAYIYYOYEAAKSPTKODYNNAKSALERALI---SGTOHDEIKKS-- 189
DB 425 SWYEGRLMAD-----YVEAKKAGKPEFYQKVLGOETVALLAKTSQTQADHILQNVG 477
QY 190 --YDKLSDY 197
DB 478 FGHKNKVSILY 487

RESULT 33
AAB14972
ID AAB14972 standard; Protein: 1115 AA.
XX
AC AAB14972;
XX
DT 01-DEC-2000 (first entry)
XX
DE N. meningitidis antigenic iron repressible protein frpa.
XX
KW Iron repressible protein; frpa; haemolysin; vaccine.
XX
OS Neisseria Meningitidis.
XX
PN US6086896-A.
XX
PD 11-JUL-2000.
XX
PF 14-OCT-1994; 94US-0323477.
XX
PR 28-JUL-1992; 92US-0920963.
XX
PR 16-JUL-1990; 90US-0552649.
XX
PA (IMCL-) IMCLONE SYSTEMS INC.
XX
PI Thompson S, Sparling PF;
XX
DR WPI; 2000-531325/48.
XX
DR N-PSDB; AAA65197.
XX
PT New antigenic iron repressible protein (frpa) isolated from Neisseria
PT meningitidis related to the hemolysin family of toxins useful for
PT identifying N.meningitidis and distinguishing it from N.gonorrhoeae
XX
PS Claim 1; Fig 3; 30pp; English.
XX
CC The present sequence is the antigenic iron repressible protein frpa
CC of Neisseria meningitidis. This sequence is derived from a clone
CC from a N.meningitidis genomic library. The clone was isolated using a
CC monoclonal antibody that was known to recognise iron-regulated proteins.
CC The frpa protein can be used as an immunogen to produce antibodies which
CC are capable of recognizing N.meningitidis and distinguishing
CC meningococcal cells from gonococcal cells in a sample and for preparing
CC vaccines against N.meningitidis.
XX
SQ Sequence 1115 AA;

Query Match 8.6%; Score 90; DB 21; Length 1115;
Best Local Similarity 23.2%; Pred. No. 4;
Matches 58; Conservative 33; Mismatches 69; Indels 90; Gaps 15;

QY 1 MKIRKMPVMVAMGLVTSACOSTPIPPKNNPOLAQITQAIISLDNGK--LDOAKQO-- 56
DB 552 MRLSTDTWTQTANEGIAL-----TP-----SQVAQLKKNALVSLSDAKAIDAARDRIA 600
QY 57 -LDAALSADROFAPAYRTLAKVQASEDA-----THOTKORLFEKAI---ELNPK 103
DB 601 VLDDYTGODSS-----TL--YFMSEEDALNIYKVTNDYDHLAKNITQNLFPOTRIOP- 651
QY 104 DMOSYMDYGFYLVOMG-----DLSGALIYED-----KPSRAI----- 135

```

```

DB 652 -----YLNQISFKMENDFTTLPDSGLVOAFNHVETNPQKAFVDLAEMLAAGELR 701
QY 136 -GYEGRVAAIEMNAYIYYOYEAAKSPTKODYNNAKSALERALI---SGTOHDEIKKS-- 189
DB 702 SWYEGRLMAD-----YVEAKKAGKPEFYQKVLGOETVALLAKTSQTQADHILQNVG 754
QY 190 --YDKLSDY 197
DB 755 FGHKNKVSILY 764

```

```

RESULT 34
AAR70982
ID AAR70982 standard; Protein: 1091 AA.
XX
AC AAR70982;
XX
DT 30-NOV-1995 (first entry)
XX
DE Megakaryocyte stem cell maturation factor clone 1204B.
XX
KW Megakaryocyte stem cell maturation factor; clone 1204B; diagnosis;
KW treatment; abnormal differentiation; megakaryocytic leukaemia.
XX
OS Homo sapiens.
XX
PN JP07067658-A.
XX
PD 14-MAR-1995.
XX
PF 03-SEP-1993; 93JP-0243767.
XX
PR 03-SEP-1993; 93JP-0243767.
XX
PA (MOCH ) MOCHIDA PHARM CO LTD.
XX
DR WPI; 1995-143852/19.
XX
DR N-PSDB; AAQ86236.
XX
PT Novel DNA involved in differentiation of stem cells to
PT mega:karyocyte(s) - used in diagnosis and treatment of diseases
PT involving abnormal differentiation, e.g. megakaryocytic leukaemia
XX
PS Claim 6; Pages 18-23; 41pp; Japanese.
XX
CC AAQ86236 encodes AAR70982 the megakaryocyte (MKC) stem cell maturation
CC factor clone 1204B. The DNA is involved in the differentiation of
CC stem cells to MKCs and platelets, it can be used as a probe to
CC diagnose diseases involving abnormal differentiation, it may also
CC be used in the treatment of MKC leukaemia. Antibodies raised
CC against the protein can used to immunologically measure the MKC's
CC level of differentiation.
XX
SQ Sequence 1091 AA;

Query Match 8.4%; Score 87.5; DB 16; Length 1091;
Best Local Similarity 21.0%; Pred. No. 7;
Matches 37; Conservative 36; Mismatches 74; Indels 29; Gaps 6;

QY 48 GKLDOAKQOQLDAALSADROFAPAYRTLAKVQASEDAHQTKAQRLEFEKAIENPDMOS 107
DB 273 GDEKNSQCEKEKLVKAYPNNVETMTLGLSYAASEDOEKRDIAKGLKVTGEQYPDDEVA 332
QY 108 YMDYGFYLVOMGLSGALIYFDKPSRAIGYEGRVAA-----IENNAYIYH----- 153
DB 333 WIELA-QILLEQDTIOGALSAYGTATRII--QEKVQADVPEILINNAGALHFRIGNIGEA 389
QY 154 -----GYEAKSPTKD--YNNAKSA-----LERALISGTOHDEIKSYDKLSDY 197
DB 390 KYFLASIDRAKAEAEHDEHYNAIVSYTSYNLARLYEAMCEPHEAKLYKNILREH 445

```

RESULT 35
AAR70983 standard; Protein; 1091 AA.
XX ID AAR70983 standard; Protein; 1091 AA.
XX AC AAR70983;
XX DT 30-NOV-1995 (first entry)
XX DE Megakaryocyte stem cell maturation factor clone 1204X.
XX KM Megakaryocyte stem cell maturation factor; clone 1204X; diagnosis;
XX treatment; abnormal differentiation; megakaryocytic leukaemia.
XX OS Homo sapiens.
XX PN JF07067658-A.
XX PD 14-MAR-1995.
XX PF 03-SEP-1993; 93JP-0243767.
XX PR 03-SEP-1993; 93JP-0243767.
XX PA (MOCH) MOCHIDA PHARM CO LTD.
XX DR WPI; 1995-143852/19.
XX DR N-PSDB; AAO86237.
XX PT Novel DNA involved in differentiation of stem cells to
XX mega:karyocyte(s) - used in diagnosis and treatment of diseases
XX involving abnormal differentiation, e.g. megakaryocytic leukaemia
XX PS Claim 6; Pages 23-27; 41pp; Japanese.
XX CC AAO86237 encodes AAR70983 the megakaryocyte (MNC) stem cell maturation
XX factor clone 1204B. The DNA is involved in the differentiation of
XX stem cells to MNCs and platelets, it can be used as a probe to
XX diagnose diseases involving abnormal differentiation, it may also
XX be used in the treatment of MNC leukaemia. Antibodies raised
XX against the protein can used to immunologically measure the MNC's
XX level of differentiation.
XX SQ Sequence 1091 AA;
Query Match 8.4%; Score 87.5; DB 16; Length 1091;
Best Local Similarity 21.0%; Pred. No. 7;
Matches 37; Conservative 36; Mismatches 74; Indels 29; Gaps 6;
QY 48 GKLDOAKQOLDAALASADROPAPAYRTLAKYVQASEDAHOTKQRLFEKAIEINPKDMOS 107
DB 273 GDKENASQCEFEKVLKAVPNNYETMKILGSLYASSEDEKRDIAKHKKVTEQYPPDVEA 332
QY 108 YMDYGFVLVOMGDISGLITFDKPSRAIGYGRVVA-----IENNAVITYH----- 153
DB 333 WIELA-QILQCTDIOGALSAYGTATRL--QEKVQADVPEIINNVGALHFRIGNLGEAK 389
QY 154 -----QYEAKSPTKDD--YNNAKSA-----LEBALSGHODEIKSYKLSIDY 197
DB 390 KYFLASLDRAKAEKAEHDEHYNAISVTTSTNLAKLYAMCEPHEAKLYNNILREH 445
RESULT 36
AAU23157 standard; Protein; 160 AA.
XX ID AAU23157 standard; Protein; 160 AA.
XX AC AAU23157;
XX DT 17-DEC-2001 (first entry)
XX DE Novel human enzyme polypeptide #243.
XX KM Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX ligase; hyperproliferative disorder; immunodeficiency disorder;
KW

KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX OS Homo sapiens.
XX PN WO20015301-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01239.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214686.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220263.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226581.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.

PN W09732980-A1.
 XX
 PD 12-SEP-1997.
 XX
 PF 07-MAR-1997; 97MO-CA00163.
 XX
 PR 03-JAN-1997; 97US-0778570.
 PR 08-MAR-1996; 96US-0613009.
 XX
 PA (CONN-) CONNAUGHT LAB. LTD.
 XX
 PI Du R, Harkness RE, Klein MH, Loosmore SM, Myers LE,
 PI Schryver AB, Yang Y;
 DR WPI; 1997-457533/42.
 DR N-PSDB; AAT95251.
 XX
 PT DNA encoding transferrin receptor of a Moraxella strain - also
 PT proteins, useful in vaccines, as diagnostic agents and in the
 PT production of antibodies
 XX
 PS Claim 6; Fig 27; 162pp; English.
 XX
 CC The present sequence is the Moraxella catarrhalis R1
 CC transferrin binding protein tbpb, which can be used as an
 CC immunogen, e.g. in vaccines to protect against diseases caused
 CC by M. catarrhalis (specifically otitis media), or to raise
 CC antibodies for diagnosis and therapy. It can also be used as a
 CC carrier for other antigenic determinants, e.g. of bacteria
 CC containing polysaccharide antigens or abnormal polysaccharides
 CC present on tumour cells, particularly to make conjugate vaccines.
 CC The tbpb DNA can be used to detect nucleic acid encoding
 CC transferrin receptor protein, e.g. for diagnosis or gene
 CC isolation, by usual hybridisation assays.
 XX
 SQ Sequence 714 AA;
 XX
 Query Match 8.2%; Score 86; DB 18; Length 714;
 Best Local Similarity 22.8%; Pred. No. 5.5;
 Matches 52; Conservative 25; Mismatches 75; Indels 72; Gaps 13;
 QY 11 MANGVLVSAC-----GSTPIPKNNPOLAIFQIAISLDMGK-LDQAKQQLDAALSA 63
 DB 100 MGYGMAISKINLYDQOQTPIDAKN-----ITLDGKKQVADNQGKSLPFLSDV 147
 QY 64 DRGFAPARYRLAKYQASDA-----THQTK-----AQRLEF----- 95
 DB 148 ENKLIDGY-IAGKNEADKNAIGERIKRENEQNKKISDEELAKKIKENVRKSPFQVLS 205
 QY 96 ----KATELNPKMOS-----YMDYGFYLVMQDLSGALIFPKPRAIGYEGRYVAIE 145
 DB 206 SIAKTHSNDKTKATTRBLKAYVDYGYLV--NDANYLVTKDKPK--LMSGPV----- 257
 QY 146 NMAYIYYHYQYEAKS-PTKD--DYNNAKSALERALISGTQHDEIKSY 190
 DB 258 --GGVFYNGSTYAKELPTQDAVKYKGWDMFTDVAKKRNFSEVKERY 303
 XX
 RESULT 38
 ABB62022
 ID ABB62022 standard; Protein; 1238 AA.
 XX
 AC ABB62022;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 12858.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX

PN W0200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001MO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL06125.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 12858; 21dp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1238 AA;
 XX
 Query Match 8.2%; Score 86; DB 22; Length 1238;
 Best Local Similarity 25.4%; Pred. No. 12;
 Matches 50; Conservative 25; Mismatches 80; Indels 42; Gaps 9;
 QY 13 MGVLVSACSTPIPKNNPQ--LAQIRQIAISLDMGKLDQAKQQLD-----AALSADR 65
 DB 931 VGSHTSAVSKSIEAKNERONILQAKTDCIVPLRGLSLDPAVRQSDPDVSTSAAMN 990
 QY 66 QFAPARYRLAKYQASDAATHQTKAQLFEKALINPKMOSYMDYGFYLVMQDLSGAL 125
 DB 991 IIEVDYSLPREYTKLKD-----DSARKKTHEMLQKDLQSLDV-LERIQTPNPK-AL 1041
 QY 126 IYFDKPSRAIGYEGRYVAIENMAYIYYHYQYEAKSPTKDDYNNAKSALERA-----LI 178
 DB 1042 QKLD-----AVTEKYQSTNEEFNAKKAK-----RAKAAFERVAKNERRSRV 1084
 QY 179 SGTQH--DEIKKSYDKL 193
 DB 1085 ACCQHTSDAIDGIYKKL 1101
 XX
 RESULT 39
 AAG29689
 ID AAG29689 standard; Protein; 251 AA.
 XX
 AC AAG29689;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 35367.
 XX
 KM Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX

PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144322.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145112.
PR 23-JUL-1999; 99US-0145115.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151458.
PR 01-SEP-1999; 99US-0151930.

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.2%; Score 85.5%; DB 21; Length 265;
Best Local Similarity 22.4%; Pred. No. 1.5;
Matches 47; Conservative 34; Mismatches 70; Indels 59; Gaps 10;

QY 16 VLSACOSTPI-----PPKNP-QIAQIRTOIAISLLDMKIDOKKQOLDALSDRQFAPR 70
DB 52 IVKALQSLSDHGMPPENDPMWMSNIVEPIVDSCLD---EEDKREKSA---SQERFLEX 104
QY 71 YRTIAK-----VYASEDAHQTAORL-----FEKAI-----ELNPKMQSYV 109
DB 105 FKRVESVAORLNEQPIVAHSENTEDGSGIRRLSKKEFFDALNVAMETTRKQOGKX 164
QY 110 DYGFYVOMGDLSCALYFDKPSRAIGYEGRVVAIENMAYIYYHOYEAAKSPKDDYNN 169
DB 165 SKGILRAVLDTVA-----PSATLPPICAVSQMDMMI-----MEALKRVANGDGNV 210
QY 170 KSALERALISGTOHDEIKSYDKLSDYKL 199
DB 211 K-----EEEFKKTMAETIGSIML 228

Search completed: July 8, 2003, 11:31:23
Job time : 73 secs

ADDRESSEE: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,046
FILING DATE: 16-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00340101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 981 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-649-046-2

Query Match 10.2%; Score 106.5; DB 2; Length 981;
Best Local Similarity 23.7%; Pred. No. 0.0099;
Matches 45; Conservative 26; Mismatches 76; Indels 43; Gaps 7;

QY 33 LAQRTQTAISLDMGKIDQAKQDLALSDROPAPRYRLAKVYQASEDATHTQTAQR 92
DB 218 LAIVLTDTGTSIKLAGNTQEGIOKYEALKTIDPHYAPAYNMGVY--SEMGQYDNLAS 274
QY 93 LFEKAIENPKDMOSYMDGYFLVQMGDLSGALTYFDK-----PSRAIGYGRVAL--- 144
DB 275 CYEKALALERPMTABRYCMAGVITYKRGDLEWATTCYERCLAVSPNFEIAKNNMALITLDL 334
QY 145 -----ENNAAY-----IYHQYEAKSPITXDYNNAKSALERALLISGTHDEIKKS 189
DB 335 GTKYKLEGDVYQGVAYYKALYYNWHVADAM-----YN-----LGVANGEWLK- 377
QY 190 YDKLSDPYKL 199
DB 378 FDMALVFEYL 387

RESULT 3
US-08-323-477-2
Sequence 2, Application US/08323477
Patent No. 6086896
GENERAL INFORMATION:
APPLICANT: Sparling, P. Frederick
APPLICANT: Thompson, Stuart
TITLE OF INVENTION: ANTIGENIC IRON REPRESSIBLE PROTEINS FROM
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Imclone Systems Incorporated
STREET: 180 Varick Street
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,477
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/920,963
FILING DATE: 28-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Felt, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: SPA-2-2P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-645-1405
TELEFAX: 212-645-2054
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1115 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-323-477-2

Query Match 8.6%; Score 90; DB 3; Length 1115;
Best Local Similarity 23.2%; Pred. No. 0.7;
Matches 58; Conservative 33; Mismatches 69; Indels 90; Gaps 15;

QY 1 MKIRVKPMVMANGVLVLSACQSTPIPPKNNPOLAQIRTOIAISLDMGK--LDAQKQ-- 56
DB 552 MRLSTDMTQYANEGIAL-----TP-----SQVAQKKNAVLSDKAAIDPAARDRIA 600
QY 57 -LPAALSARQRFAPARTLAKVYQASEDA-----THQTAQQLFEKAT---ELNKR 103
DB 601 VLDPYVQDSS-----TL--YMSSEDAINIVKVTNDYDHLAKNIYONLLFQTRLP- 651
QY 104 DMQSYMDGYFLVQMG-----DLSGALTYFD-----KPSRAI----- 135
DB 652 -----YLNQISFKKENDFTFLDFSGLVQAFHNVETNPQKAFVLAEMLAYGEIR 701
QY 136 -GYEGRVAALENMAYIYHQYEAKSPITXDYNNAKSALERALLISGTHDEIKKS-- 189
DB 702 SWYEGRRIMAD-----YVEAKKAKGFEDYQVKGQFTVALLAKTSQTQADDILQNVG 754
QY 190 --YDKLSDPY 197
DB 755 FGHKNVSLY 764

RESULT 4
US-09-059-584-49
Sequence 49, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Harkness, Robin E
APPLICANT: Loosmore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 713 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-059-584-49

Query Match      8.2%; Score 86; DB 4; Length 713;
Best Local Similarity 22.8%; Pred. No. 0.99;
Matches 52; Conservative 29; Mismatches 75; Indels 72; Gaps 13;

QY 11 MAMGVLSAC-----GSTPIPPKNNPQLAQIRFOIAISLDMGK--LDQAKQQLDAALSA 63
DB 100 MGVMGLSKINLYDQDTPLDANK-----ITLIDGKKQVADNQSPLPSLDV 147
QY 64 DRQFAPAYRTLAKYQVASEDA-----THQTK-----AORLFE----- 95
DB 148 ENKLLDGY--IAIGNENDKKAIGERIGRENGKKSIDSEELAKIKENYKSPFOQVLS 205
QY 96 ----KAIELNPKCMQOS-----YMDYGFYLQWGDLSGALIYFDKPSRAIGYGRVAIE 145
DB 206 SIAKTFHSHNDKTKATRTPLKTVYDGYLV--NDANYLVTKTDKPK--LMNSGPV----- 257
QY 146 NMAVIYVHOYEAKS--PTKD--DYNNAKSLERALLSGTHDHIKSY 190
DB 258 --GGVFYNGSTTAKELPTDQAVYKKGHMDPTVAKKRNFSSEKERY 303

RESULT 5
US-08-471-119A-2
Sequence 2, Application US/08471119A
Patent No. 5827706
GENERAL INFORMATION:
APPLICANT: Leitner, Ernst
APPLICANT: Schneider, Elisabeth
APPLICANT: Schoergerdorfer, Kurt
APPLICANT: Weber, Gerhard
TITLE OF INVENTION: Cyclosporin Synthetase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5827706artis Corporation
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,119A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kaesenoft, Melvyn
```

```
REGISTRATION NUMBER: 26,389
REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 503 8474
TELEFAX: 201 503 8807
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 15281 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Toxoplasma gondii
STRAIN: ATCC 34921
US-08-471-119A-2

Query Match      8.2%; Score 85.5; DB 2; Length 15281;
Best Local Similarity 26.6%; Pred. No. 91;
Matches 34; Conservative 23; Mismatches 48; Indels 23; Gaps 6;

QY 23 TPIPPKNNPQLAQI-----RTQ-IAISLDMGKLDQAKQQLDAALSA--DRQFAPAYRTL 74
DB 4367 TPIANNRDELEQIGFVNTQIRITVNEDEFESLVQVNSTATAPAHQDVPEKIV 4426
QY 75 AKYQVASEDAITHTQKQRLF-----EKA-IELNPKDMSYMDYGFYLQWMD 120
DB 4427 STLLPSSRDASRNPLVQMFVAHVSQKIGLKNLHNSHVPLTITRDLFEHLFQOD 4486
QY 121 -LSGALIV 127
DB 4487 KLEGSILY 4494

RESULT 6
US-09-134-001C-3811
Sequence 3811, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3811
LENGTH: 330
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3811

Query Match      8.1%; Score 84.5; DB 4; Length 330;
Best Local Similarity 18.5%; Pred. No. 0.48;
Matches 56; Conservative 40; Mismatches 75; Indels 131; Gaps 13;

QY 10 VMAMGVLSACGSTRPIPPKNNPQLAQIRFOIAIS--LDMGK----- 49
DB 16 VTRASALLDAGCGSNATESKONTLISKAGDVADVAKKQKGEQIANTSFSIVLNKRYLAD 75
QY 50 -----LDQAKQ-----LDAA 60
DB 76 KYEDKXVTDIDDKIDKEKQYGGKQGFESMLKQGSMSLDYKEQKKLSAYQKQLDDK 135
QY 61 LSAHQF-----APAYFTLAKYQAS-----EDATHTQTAQRLFEKAIELNPKDMSYMD 110
DB 136 NVSDKEIKENSKTSHILIKYKSSSDKEGLSDKAKAKAEKIK-QUEVEKNPV----- 187
```

QY 111 YGFLVQMGDLS-----GALIYEDKPSRAIG-----EGRVVAIINMAY 149
Db 188 -----KFGELAKESDSSAKKDSGLGVYIGQVNDSEKALFKLKGEYSKVXTDY 241
QY 150 IYHGYEAASPTKDDYNNAKSALERALISGTQDEIK-----SYDKLSYDKLSD 202
Db 242 -----GHYIIRKADKETDFNSEKSNIKOKLI-----EENYQKKPKLLTAYAKELKEYKV--D 291
QY 203 YK 204
Db 292 YK 293

RESULT 7
US-09-059-584-46
; Sequence 46, Application US/09059584
; Patent No. 6440701
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,584
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/778,570
; FILING DATE: 03-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 706 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-059-584-46

Query Match 8.0%; Score 83.5; DB 4; Length 706;
Best Local Similarity 21.8%; Pred. No. 1.8;
Matches 52; Conservative 36; Mismatches 84; Indels 67; Gaps 13;

QY 11 MAMGLVVSAC-----OSTPIPEKN-----NPOAQ-----IRTOIALSLD--MGKL 50
Db 100 MGGMALSKNMLHQQDPTDEKQITLDEKKQVAKGEKSPPLPESLDVENKLDGYIAKK 159
QY 51 DQA-----KQDLAALSADRPAPAYRTIAKYQASEDATQYTKQRLF----- 94
Db 160 NEADKNAIGRIKKNDKSLK-----AEIAK--QIKEDVRKSHEQVLSLKKK 209

QY 95 -----EKALIELNPKDMOSYMDGYFLVQMGDLSGALIYFPK-----PSRAIGEGRVVAI 144
Db 210 IFHSDGTTKATFTDLO-YVDGYVLVNDGY--LTKTBEIMNLGFGVGVNNGTTAK 266
QY 145 E-----MAYIYHGYEAASPTKDDYNNAKSALERALISGTQDEIKSYDKLSDYK 198
Db 267 ELFTQDAVKYKGHWDMTDAKQNRPFSEYKENILOAGRYYGAS--KDEYNRLLTDEK 322

RESULT 8
US-07-689-008-2
; Sequence 2, Application US/07689008
; Patent No. 5268274
; GENERAL INFORMATION:
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Calhoon, Roger D
; APPLICANT: Fear, Anna L
; APPLICANT: Gelfand, David H
; APPLICANT: Meade, James H
; APPLICANT: Tal, Ronny
; APPLICANT: Wong, Hing
; APPLICANT: Benzman, Moshe
; TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE
; EXPRESSION OF CELLULOSE SYNTHASE OPERON
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McCutchen, Doyle, Brown & Enersen
; STREET: Three Embarcadero Center
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/689,008
; FILING DATE: 19910422
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 337,194
; FILING DATE: 12-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 496,236
; FILING DATE: 23-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy, Lisabeth Felix
; REGISTRATION NUMBER: 31547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 393-2000
; TELEFAX: (415) 393-2286
; TELEEX: 340817 MACPAG SFO
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3031 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-689-008-2

Query Match 8.0%; Score 83.5; DB 1; Length 3031;
Best Local Similarity 23.0%; Pred. No. 15;
Matches 31; Conservative 20; Mismatches 57; Indels 27; Gaps 4;

QY 26 PPKNNPQAAQIRTOAISLDMGKLDQAKQDLAALSADRPAPAYRTIAKYQASDAT 85
Db 1839 PGGFPDAGLARGAGYQULNAGRLAAHDSFQSALQINSHDADSLGMLGVNSRQDGT- 1897
QY 86 HQYAKQRLFEKALIELNPKDMOSYMDGYFLVQMGDLSGALIYFDPKPSRA-IGYEGRVVAI 144
Db 1839 PGGFPDAGLARGAGYQULNAGRLAAHDSFQSALQINSHDADSLGMLGVNSRQDGT- 1897

Db 1898 --A6ARRYFEANMAADPKTADRW-----RPAAGNAVSGEYASV 1934
QY 145 ENMAYIYYHOYEAK 159
Db 1935 ROL--IAAHOYTEAK 1947

RESULT 9
US-07-618-946B-22
; Sequence 22, Application US/07618946B
; Patent No. 5453371
; GENERAL INFORMATION:
; APPLICANT: YUJI SHIBANO et al.
; TITLE OF INVENTION: Bacterial Collagenase Gene
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoch, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 KB
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618,946B
; FILING DATE: 19901127
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX: 440293 WNDP UI
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 739 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLER:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:

OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-07-618-946B-22

Query Match 7.9%; Score 82; DB 1; Length 739;
Best Local Similarity 26.2%; Pred. No. 2.8; 57; Indels 38; Gaps 11;
Matches 44; Conservative 29; Mismatches

QY 30 NPQLAQI---RTQIAISLDMGKLDQAKOQDPAALSADROF--APAYRTLAKYQASEDA 84
Db 178 NEQFVQIIINGQTDLAALGDF-----LRASSIGAEDEFMANNRGRELRLTKYTGNA 230
QY 85 THQTKAQ--RLFEK-----AIEINPKDMOSY---MDYGFYLVQMGDLGALI---Y 127
Db 231 SSVVKSQLSRIFRQYEMVGRGDAVWLAADTASYADCESEFGICNFET-ELKGLVLVSQTV 289
QY 128 FDEP-----SRAGYGRVVAIENMAYI--YHQ-YEAKSPFKDDYN 167
Db 290 TCSPTIRILSQNTQBOHAAACSKMGVEGYFHSLETGPVQVDHN 337

RESULT 10
US-07-618-946B-23
; Sequence 23, Application US/07618946B
; Patent No. 5453371
; GENERAL INFORMATION:
; APPLICANT: YUJI SHIBANO et al.
; TITLE OF INVENTION: Bacterial Collagenase Gene
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoch, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 KB
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618,946B
; FILING DATE: 19901127
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX: 440293 WNDP UI
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 814 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear

```

MOLECULE TYPE:
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-07-618-946B-23

```

```

Query Match 7.9%; Score 82; DB 1; Length 814;
Best Local Similarity 26.2%; Pred. No. 3.2;
Matches 44; Conservative 29; Mismatches 57; Indels 38; Gaps 11;

QY 30 NPOLAQI---RQIQAISLDMGKLDQAQQLDAALSADROF--APAYRTIAKVQASEDA 84
DB 253 NEQVQIIGNGDTDLAKALGDFR-----LRASISGAEDFEWAAGSELGRITVYTGNA 305
QY 85 THQTKAQ--RLFEK-----AIELNPKDMQSY---MDYGFYLVQMGDLSGALII---Y 127
DB 306 SSVKSSQSLRFPEQYEWYGRGDAVLAADTASYYADSCSEFGICNFET-ELKGLVLSQTY 364
QY 128 FDKP-----SPAIGEGRVVAIENNAIY--YYHQ-YEAKSPTKDDYN 167
DB 365 TCSPTIRILSONMTQEOHAACSKMGYEGYFHQSLTGEQPVKDDHN 412

RESULT 11
US-09-147-236-5
Sequence 5, Application US/09147236A
Patent No. 6316251
GENERAL INFORMATION:
APPLICANT: TONOUCHI, Naoto
APPLICANT: TSUCHIDA, Takayasu
APPLICANT: YOSHINAGA, Fumihito
APPLICANT: TAHARA, Naoki
APPLICANT: HAYASHI, Takahisa
TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE
FILE REFERENCE: 6537-011-OPCT
CURRENT FILING DATE: 1999-04-08
CURRENT APPLICATION NUMBER: US/09/147,236A
EARLIER APPLICATION NUMBER: PCT/JP97/03633
EARLIER FILING DATE: 1997-10-09

```

```

NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1326
TYPE: PRT
ORGANISM: Acetobacter xylinum
FEATURE:
OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, t, or
US-09-147-236-5

```

```

Query Match 7.7%; Score 80.5; DB 4; Length 1326;
Best Local Similarity 20.9%; Pred. No. 9.4;
Matches 31; Conservative 17; Mismatches 47; Indels 53; Gaps 4;

QY 26 PPKNNPQIAQIRIQIQAISLDMGKLDQAQQLDAALSADROFAPAYRTIAKVQASEDA 85
DB 283 PPGPPDPKQIARQAGYQQLNAGRLSAAEQS----- 313
QY 86 HQTKAQRLEPKAIELNPKDMQSYMDYGFYLVQMGDLSGALITPD-----KPS 132
DB 314 -----FQSAIQINSHDADSLGGMGLVSMRGDTREAHPIYEEAAMADPKTRADRRPA 365
QY 133 RA-IGYEGRVVAIENNAIYIYHQYEAK 159
DB 366 LAGMAVSGDYAAVRQL--IAHQYTEAK 391

```

```

RESULT 12
US-09-051-019-2
Sequence 2, Application US/09051019
Patent No. 6103229
GENERAL INFORMATION:
APPLICANT: KAHMANN, Regine and QUADBECK-SEEGER, Claudia
TITLE OF INVENTION: Regulatory gene from Ustilago maydis
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Keil & Weinlauf
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
COMPUTER: IBM AT-compatible, Pentium processor
OPERATING SYSTEM: Windows 98
SOFTWARE: wordperfect version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051,019
FILING DATE: 31-MAR-1998
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2289 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-051-019-2

```

```

Query Match 7.7%; Score 80.5; DB 3; Length 2289;
Best Local Similarity 23.1%; Pred. No. 21;
Matches 39; Conservative 27; Mismatches 48; Indels 55; Gaps 7;

QY 54 KQQLDAAL-----SADROFAPAYRTIAKV-----YQASEDAHQTKAQRLE-----F 94
DB 1099 KQSLDASVDADTESTDSRPAIYALIGELSLHFDAPETASITMAQELDEFGRCDREV 1158
QY 95 EKAIELNPKDMQSYMDYGFYLVQMGDLSGALITPD-----VOMGDLGALITFDKPSRAIGYEGRVVAIENN-- 147
DB 1159 QQGGERTNLKQCESILITIGSSLNVADAPQIKELS-----DYVERRKWIQEVTES 1205
QY 148 --AYIYHQYEAKSPTKDDYNNAKSALEALALISGYOHDIKSYDRL 194

```


Db 1206 FDTLYVYHE-----VAEILDRADSCGLQDHRLRNLEQL 1240

RESULT 13
US-09-134-001C-3159
Sequence 3159, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3159
LENGTH: 10182
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 7.7%; Score 80; DB 4; Length 10182;
Best Local Similarity 19.9%; Pred. No. 2e+02;
Matches 43; Conservative 36; Mismatches 71; Indels 66; Gaps 8;

QY 21 QSTPIPKNNPOLAQIRTOI-----AISLI-----DMGLDQAKQQLDALS-- 62
DB 3804 EANSILTNBNPOVSQVTAALNIKAVQPELDKAIAMKNNKNNALVQAKQQLQOIYNEV 3863
QY 63 -----ADRFAPAYRTLAKVYQASEDAHQTAQSLFEKATLELN 102
DB 3864 DPLQMGTTDPAANNYKSKREHEDETQKAOQIINNDATEQOITNETNRVNOALINAKK 3923
QY 103 KDM---QSYNDYGF-YLVQMGDLGALIFDPKPSRAIGEGRVAIEMMAYIYYHYEAA 158
DB 3924 NDLRADSKQLENNAYNOIIOQVDTNG-----KKPKS-----IQYQAA 3960
QY 159 KSPTKDDYNNAKSALERAL-ISTGQDEIKKSYDKL 193
DB 3961 ROALETQYNNAKSEAHQILENSNPSVNEVAALQKV 3996

RESULT 14
US-09-059-584-57
Sequence 57, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Harkness, Robin E
APPLICANT: Loosmore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 512 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-059-584-57

Query Match 7.6%; Score 79.5; DB 4; Length 512;
Best Local Similarity 22.9%; Pred. No. 3.1;
Matches 41; Conservative 21; Mismatches 62; Indels 55; Gaps 6;

QY 25 IPPKNNPOLAQI-----RTQIATSLDMGKLDQAKQQLDALSADRFAPAYRTLAKVQ 79
DB 78 LTPKTEPILAQITNPAPEDTLIDAEALLDNPDLNHALNSAVTNNM--AGVHALPIYQ 135
QY 80 ASEDATHQTKAQRLFEKATLELNPKDMQSYNDYGFYLVQMGDLGALIFDPK--PSRAIGY 137
DB 136 KL-----PKDHQNGILGY-----ANALVALDKGNAKAIGE 167
QY 138 EGRVVAIEMMAYIYYHYEAAKSPYKDDYNNAKSALERALISTGQDEIKKSYDKLSD 196
DB 168 LRRRIAI-----MPEYVVFHLARALFMDKQNEALDQFNKLHAD 208

RESULT 15
US-08-795-475-6
Sequence 6, Application US/08795475
Patent No. 5965390
GENERAL INFORMATION:
APPLICANT: Bjvrck, Lars
APPLICANT: Sjvrck, Ulf
TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,475
FILING DATE: 11-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 100084,402D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 443 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-475-6

Query Match 7.5%; Score 78.5; DB 2; Length 443;
Best Local Similarity 27.1%; Pred. No. 3.2;
Matches 52; Conservative 23; Mismatches 66; Indels 51; Gaps 11;

QY NNPOLAOIRQIAISLDMGKLDQAQKQDAAALS-ADROFAPARTLAKVYQASEDATHQ 87
DB NNPOLAOIRQLR-----HENKDLKARLENAMVAGDFKRA-ELEKAKQALD----- 62
QY 88 TKAQRLFEKALINPKDMQSYMDYGFYLVQMGDLGALITYDKPSRAIGYGRVVAIE-- 145
DB 63 ---QR---KOLETKELKELQ---QDY-----DLAKESTSWDRQRLKELEKEKELELA 106
QY 146 -NNAYIYHQ-----YEAKSPTKDDYNNAKSALBRALISQTHDEIKSY 190
DB 107 IDQASRDYHATVALEKELEKKALELAIDQASQDYNRA-NVLEKELETTTREGEDINRN- 164
QY 191 DKLSDYKLSQ 202
DB 165 --LIGNAKLELD 174

RESULT 16

US-08-985-343-6
Sequence 6, Application US/08985343
Patent No. 6468782

GENERAL INFORMATION:

APPLICANT: Tumacacife, Alan G.
APPLICANT: Welsh, David T.
APPLICANT: Rose, Bruce J.
APPLICANT: Dhaliwal, Kamaljit S.
APPLICANT: Colaco, Camillo
TITLE OF INVENTION: METHODS OF PRESERVING PROKARYOTIC CELLS
TITLE OF INVENTION: AND COMPOSITIONS OBTAINED THEREBY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CA
COUNTRY: USA
ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985.343

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Lehmanard, Susan K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 26374-20017.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 813-5600
TELEFAX: (650) 494-0792
TELEK: 706141 MRSNPOERS SFO
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 473 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-985-343-6

Query Match

7.5%; Score 78.5; DB 4; Length 473;

Best Local Similarity 22.7%; Pred. No. 3.5;
Matches 47; Conservative 32; Mismatches 65; Indels 63; Gaps 12;

QY 24 PIPKXNPOLAOIRFOI-AISLDMGKLDQAQKQDAAALSAD-----RQ 66
DB 239 PLPPK-----LAOLKALKNVQNIIFSVRLDYKGLPERFLAYALELLEKYPQHGIKIRYQ 294
QY 67 FAPAVRTLAKVYQASEDATHQKARLFEKALINPKDMQSYMDYG-----FYLQMG 119
DB 295 IAPTSRGDVAQO---DIRHQ-----LENEAGRINCK-----YQGLGWTPLYLINO--- 337
QY 120 DISGALITYDKP--SRAGYE--GRVVALENMAYIYHGYEAKSPTK-----DDYNN 169
DB 338 -----HFDRKLIMKIFRYSVDGLVTPLRDGMNLVAKRYVAQDPANGLVLSQFAGA 390
QY 170 KSALEBALI-SGTQHDIEIKSYDKLS 195
DB 391 ANELTSALIVNPYDRDEVAALDRALT 417

RESULT 17

US-08-274-121B-4
Sequence 4, Application US/08274121B
Patent No. 6133034

GENERAL INFORMATION:

APPLICANT: Arne Reider Strom
APPLICANT: Inga Kaasen
APPLICANT: Olaf Bay Styrvold
APPLICANT: John McDougall
TITLE OF INVENTION: Methods and Compositions
TITLE OF INVENTION: Related To The Production
TITLE OF INVENTION: Of Trehalose
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Filth Street
CITY: Davis
STATE: CA
COUNTRY: USA
ZIP: 95616

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: Microsoft Word 5.1(a)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/274.121B
FILING DATE: 12-JULY-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/893,099
FILING DATE: 27-MAY-1992

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Elizabeth Lassen
REGISTRATION NUMBER: 31,845
NAME: Donna E. Scherer
REGISTRATION NUMBER: 34,719
REFERENCE/DOCKET NUMBER: CGNE 86(1)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313
TELEFAX: (916) 753-1510
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 475 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-274-121B-4

Query Match 7.5%; Score 78.5; DB 4; Length 475;

Best Local Similarity 22.7%; Pred. No. 3.5;
Matches 47; Conservative 32; Mismatches 65; Indels 63; Gaps 12;

RESULT 20

US-09-074-658-19

Sequence 19, Application US/09074658

Patent No. 6184371

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M

APPLICANT: Run-Pan Du

APPLICANT: Qiljun Wang

APPLICANT: Yang, Yan-Ping

APPLICANT: Klein, Michel H

TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

NUMBER OF SEQUENCES: 78

CORRESPONDENCE ADDRESS:

ADDRESS: Sim & Mcburney

STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/074,658

FILING DATE: 08-MAY-1998

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-795

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1153

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 1076 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-074-658-19

Query Match 7.5%; Score 78.5; DB 4; Length 1076;

Best Local Similarity 21.0%; Pred. No. 11;

Matches 44; Conservative 29; Mismatches 80; Indels 57; Gaps 7;

QY 28 KNNPOLAQRTOIALSLDMKLT-----DQAKQ-----LDAALSADROF 67

DB 5 KNNKSKSKOVLKSLSLGILNTTQVALANTTDKAEADTKNLVVVLDDETVYAKN 64

QY 68 APAYRT-----LAKYQASEDATHQTAORLFEKALINPDKMSYMDYGFYLVONG-- 119

DB 65 APVSRKANEVGLGKVVKTAE-----TINKEVILNIRDLTRY-DPGIAVVEQGRG 113

QY 120 -----DLSGLIIFDKPSRAIGYGRVV-----AIEMAIYVHQVEAASP 161

DB 114 ASSGYSIRGMQNRVAVLVLDGINOAHYOGPVAGKNVAAAGALINEIENVRSVEISKGA 173

QY 162 TKDYNNAKSALERALISGTQHDEIKKSYD 191

DB 174 NSSEYSGALSGSVAFVTKTADITKDKD 203

RESULT 21

US-09-047-288-2

Sequence 2, Application US/09047288

Patent No. 6350656

GENERAL INFORMATION:

APPLICANT: KUMAR, Sanjay

TITLE OF INVENTION: DRUG BINDING PROTEIN

FILE REFERENCE: GP-50000-D1

CURRENT APPLICATION NUMBER: US/09/047,288

CURRENT FILING DATE: 1998-03-24

PRIOR APPLICATION NUMBER: 08/802,191

PRIOR FILING DATE: 1997-02-18

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 364

TYPE: PRT

ORGANISM: HOMO SAPIENS

US-09-047-288-2

Query Match 7.4%; Score 77.5; DB 4; Length 364;

Best Local Similarity 25.2%; Pred. No. 3.1;

Matches 39; Conservative 28; Mismatches 43; Indels 45; Gaps 10;

QY 50 LDOAKQOLDALASADROFAPAYRTLAKYQASEDATHQTAORLFEKALINLP-KOMOST 108

DB 229 IDQTKRIMEV-----GTSPSEVLAKI--SSEHA-----RTYQSLPMPQKDLSS- 272

QY 109 MDYGFYLVONG-----DLSGLIIFDKPSRAIGYGRVVAIEMAIYVHQV-EAKSP 161

DB 273 -----IFRGANPLAIDLGRMLVLD-----DQVSAEALAHAFSQYHDEDEB 318

QY 162 TKDYNNAKSALERALISGTQHDEIKK-SYDKLLS 195

DB 319 EAEPTDESVAKERTL-----EWEKELTYQEVLS 347

RESULT 22

US-08-802-191-2

Sequence 2, Application US/08802191

Patent No. 6376214

GENERAL INFORMATION:

APPLICANT: KUMAR, SANJAY

TITLE OF INVENTION: DRUG BINDING PROTEIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/802,191

FILING DATE: 18-FEB-1997

CLASSIFICATION: 514

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: WILLIAM T. HAN

REGISTRATION NUMBER: 34,344

REFERENCE/DOCKET NUMBER: GP50000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219

TELEFAX: 610-270-4026

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 364 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-802-191-2

PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3481
LENGTH: 415
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3481

Query Match 7.4%; Score 77; DB 4; Length 415;
Best Local Similarity 18.0%; Pred. No. 4.2;
Matches 42; Conservative 47; Mismatches 100; Indels 44; Gaps 9;

QY 3 IRVKNPVWANGVLVLSACOSTPIPKNNPQIAQRTQIAISLD----MGKLDQAKQOLD 58
DB 31 IMLAFTLIALGLFENSAHADSLPQJNGAN--OKTKVTVSNKVPAPVRLAEQVYLSR 88
QY 59 AAL--SADROFAPARLTLA---KYYQASEDA-----THQTKAQLFEKALIELNPKDMQ 106
DB 89 VALLDKSNHKAATST-TLGEPPKIYKFNKESDGNYYVVLNKKGDVIVVTTISPNPSNSK 147
QY 107 SYMDYGFYVQMGDISGALI-----YFKPSRAIGYGRVVAI-----EN 146
DB 148 ASKQNNYSINVSPLSKILNQYKQKITLILNTRKGF-----ALTEGKTVLVLKTPRN 202
QY 147 MAYIYHOYEAKSPTKDDYNNAKSALERALISGTQDEIKSYDKLSDYKL 199
DB 203 NEKTYENATSTKPKDLNDFKQTASVTKPTLEYOSTRNMVAYEVNQLKNFRI 255

RESULT 27
US-08-687-956A-23
Sequence 23, Application US/08687956A

GENERAL INFORMATION:
PATENT NO. 5861157
APPLICANT: BURNIE, JAMES P
APPLICANT: MATTHEWS, RUTH C
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
TITLE OF INVENTION: INFECTIONS DUE TO STREPTOCOCCI AND ENTEROCOCCI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO, LLP
STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH
FLOOR
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,956A
FILING DATE: 29-JUL-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9401689.6
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 50885/222892
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/861-3000
TELEFAX: 202/822-0944
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Streptococcus sobrinus
US-08-687-956A-23

Query Match 7.4%; Score 77; DB 2; Length 1566;
Best Local Similarity 20.4%; Pred. No. 28;
Matches 42; Conservative 38; Mismatches 72; Indels 54; Gaps 7;

QY 25 IPEKNNPOLAQIRTOIAISLIMGKUDQAKQOLDALSDROFAPAY-RTLAKVYQAS-- 81
DB 194 ITNENQKADYEAKLAQYQKDLAAVOQANNDSQAAVYAAKE---AYDEKLARVQANAA 250
QY 82 ----ED--ATHQTKAQLFEKALIELNPKDMQSYMDYGFYLVQMGDISGALIYFDKPSRA 134
DB 251 AKKEVEBALAANTTTNKEQIKAEVVAIQORNAQAKADYEAKLAQ----- 293
QY 135 IGYGRVVALENMAYIYHOYEAKSPTQDYN-----AKSALERALISGT----- 181
DB 294 --YEKDLAAQSGNATNEADYQAKKAAYEQELARVQANAAQAYEQALANTAKNAOI 351
QY 182 -----QHDEIKSYDKLSDYK 198
DB 352 TAENEAIQORNAQAKADYEAKLAQYQ 377

RESULT 28

US-08-651-940-2
Sequence 2, Application US/08651940

GENERAL INFORMATION:
PATENT NO. 594885
APPLICANT: Stein, Bernd
APPLICANT: Yang, Maria X.H.
TITLE OF INVENTION: MITOGEN-ACTIVATED PROTEIN KINASE p38-2
TITLE OF INVENTION: AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,940
FILING DATE: 20-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.412
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-651-940-2

Query Match 7.3%; Score 76.5; DB 2; Length 364;

APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,171
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DIGI910, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 653 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-191-171-5

Query Match 7.3%; Score 76; DB 4; Length 653;
Best Local Similarity 19.7%; Pred. No. 10;
Matches 40; Conservative 28; Mismatches 87; Indels 48; Gaps 8;

QY 8 PMWAMGLVLSACOSTPIPEKNNPOLAQIRTOIAISLDMGKL-----DOAKQ 55
DB 6 PRAALALALASLLAAPPVAPAEAPHLYOV--DAARALWPLRRFWRSTGCPPLPHSQADQ 63
QY 56 QDAALISADROFAPARTLAKYVQASEDAHTQKARLFE-----KATLNPCKMQ 106
DB 64 YV--LSWDQQL-----NLAYVGAVPHRGIKQVRTWMLLELVTTGSGTGRGLSYNFTHLD 115
QY 107 SYMDY-----GFYLVQMGDLISGALLYFDKPSRAIGYEG-----RVAIENMAYIY 151
DB 116 GTLDLRENQLLPGFEL--MGSASGHFTDPEDKQVFEWKDLVSSLARRYIGRYGLAHVS 173
QY 152 YHOYEAKSPTKDDYNNAKSALE 174
DB 174 KWNFTWNEBDHHDFFDVSMTWQ 196

RESULT 32
US-09-385-707-5
Sequence 5, Application US/09385707
Patent No. 6238662
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,707
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,104
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DIGI910, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 653 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-385-707-5

Query Match 7.3%; Score 76; DB 4; Length 653;
Best Local Similarity 19.7%; Pred. No. 10;
Matches 40; Conservative 28; Mismatches 87; Indels 48; Gaps 8;

QY 8 PMWAMGLVLSACOSTPIPEKNNPOLAQIRTOIAISLDMGKL-----DOAKQ 55
DB 6 PRAALALALASLLAAPPVAPAEAPHLYOV--DAARALWPLRRFWRSTGCPPLPHSQADQ 63
QY 56 QDAALISADROFAPARTLAKYVQASEDAHTQKARLFE-----KATLNPCKMQ 106
DB 64 YV--LSWDQQL-----NLAYVGAVPHRGIKQVRTWMLLELVTTGSGTGRGLSYNFTHLD 115
QY 107 SYMDY-----GFYLVQMGDLISGALLYFDKPSRAIGYEG-----RVAIENMAYIY 151
DB 116 GTLDLRENQLLPGFEL--MGSASGHFTDPEDKQVFEWKDLVSSLARRYIGRYGLAHVS 173
QY 152 YHOYEAKSPTKDDYNNAKSALE 174
DB 174 KWNFTWNEBDHHDFFDVSMTWQ 196

RESULT 33

US-09-336-447A-15
; Sequence 15, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: PEDERBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Query Match 7.2%; Score 75; DB 4; Length 889;
Best Local Similarity 23.8%; Pred. No. 21;
Matches 45; Conservative 28; Mismatches 76; Indels 40; Gaps 7;
QY 28 KNNPOLAQRTOIAISLDWCK--LDOAQKQDLAALSADRFAPAYRTLAKVYQASDPA 85
DB 477 KNNADIAKNOADYQTELENDVKGELNLSGRLLDQKADIDNNIHIELAQODQDSSDI 535
QY 86 HQTKAQLF-EKALINPKDMQSYMDYGFYLV-----KDMQSYMDYGFYLVOMGDLGALIFDPKPS 132
DB 536 -KTLKKVNEEGLELSHLLDQKADLTQKIKALESNVEELL-----DLSGRLI----- 583
QY 133 RAIGYGRVAIENMAIYIYHOYEAKSPTKDY---NNAKSALERALISGTQHDEIKK 188
DB 584 ----DQKADIAONANI---ODLAAYNELQDQYAKQTEAIDALNKAASENTQIEDLA 635
QY 189 SYDKLSLDY 197
DB 636 ANELEDAY 644

RESULT 34
US-09-446-504-6
; Sequence 6, Application US/09446504
; Patent No. 6218150
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshiaki
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazuo
; APPLICANT: MURAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunobu
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/446,504
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1263
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-446-504-6

Query Match 7.2%; Score 75; DB 4; Length 1263;
Best Local Similarity 23.2%; Pred. No. 34;
Matches 29; Conservative 22; Mismatches 48; Indels 26; Gaps 3;
QY 85 THQTKAQLF--EKALINPKDMQSYMDYGFYLV-----QNG 119
DB 849 THDFEKGPLVSEQIVELKPOVILSKGAKYLLRVARFVDLLEKFGILPREYNAEKOE 908
QY 120 DLGSLIYFDKPSRAIGYGRVAIENMAIYIYHOY-EAKSPTKDDYNNAKSALERALI 178
DB 909 DLGLHVLIVGLAPRTSGIVRIIGFVDALVGYAHPTFAAKRRNCDEDSVLLLDALL 968
QY 179 SGTQH 183
DB 969 NFSRY 973

RESULT 35
US-09-712-266-6
; Sequence 6, Application US/09712266
; Patent No. 6333158
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshiaki
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazuo
; APPLICANT: MURAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunobu
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/712,266
; CURRENT FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 09/446,504
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1263
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-712-266-6

Query Match 7.2%; Score 75; DB 4; Length 1263;
Best Local Similarity 23.2%; Pred. No. 34;
Matches 29; Conservative 22; Mismatches 48; Indels 26; Gaps 3;
QY 85 THQTKAQLF--EKALINPKDMQSYMDYGFYLV-----QNG 119
DB 849 THDFEKGPLVSEQIVELKPOVILSKGAKYLLRVARFVDLLEKFGILPREYNAEKOE 908
QY 120 DLGSLIYFDKPSRAIGYGRVAIENMAIYIYHOY-EAKSPTKDDYNNAKSALERALI 178
DB 909 DLGLHVLIVGLAPRTSGIVRIIGFVDALVGYAHPTFAAKRRNCDEDSVLLLDALL 968
QY 179 SGTQH 183
DB 969 NFSRY 973

RESULT 36
US-09-091-889A-4
; Sequence 4, Application US/09091889A
; Patent No. 6355525
; GENERAL INFORMATION:
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: ISHINO, YOSHIZUMI

APPLICANT: KATO, IKUNOSHIN
TITLE OF INVENTION: NOVEL DNA POLYMERASE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/091,889A
APPLICATION NUMBER: US/09/091,889A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WEINER, MARC S.
REGISTRATION NUMBER: 32,181
REFERENCE/DOCKET NUMBER: 1422-0347P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1263 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-091-889A-4

Query Match 7.2%; Score 75; DB 4; Length 1263;
Best Local Similarity 23.2%; Pred. No. 34;
Matches 29; Conservative 22; Mismatches 48; Indels 26; Gaps 3;

QY 85 THQTAQRLLF-EKAIENPKDMOSYMDYGFYLV-----OMG 119
DB 849 THDFGKPLVSDQIVELKPPQVILSKKGYLLVAFVDDLKIFYGLPRFYNAKME 908
QY 120 DISGALIVDFKRSRAIGYGRVVALENMAYIYHYQY-EAKKSPTKDDNNAKSALERALI 178
DB 909 DIGHLVIGLAPHTSAGIVRILITGFVDLVGYAHYFPAKAKRNDGDEDSVMLLDL 968
QY 179 SGTQH 183
DB 969 NFSRY 973

RESULT 37
US-09-134-001C-3403
Sequence 3403, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3403
LENGTH: 370
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3403

Query Match 7.1%; Score 74.5; DB 4; Length 370;
Best Local Similarity 27.4%; Pred. No. 6.6;
Matches 29; Conservative 13; Mismatches 43; Indels 21; Gaps 4;

QY 67 FAPAVRTLA-----KYQASEDATHTQKAGLEKAIENLPKDMOSYMDYGFYLVQMGD 120
DB 169 FASATKAIPIVQNNNAISTPRSKQPAETTLAKAISV-----GIRGIQVDG 237
QY 121 LSGALITY---FDKPSRAIGYGRVVALENMAYIYHYQYEAASPTK 163
DB 238 MDALAVQATLEARERAVAGEPTV-ITLTLYRGPHMTAGDDPTR 282

RESULT 38
US-08-828-010-2
Sequence 2, Application US/08828010
Patent No. 5955355
GENERAL INFORMATION:
APPLICANT: Chen, Hong
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC
DISORDERS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,010
FILING DATE: 27-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-068
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9030
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 300 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-828-010-2

Query Match 7.1%; Score 74; DB 2; Length 300;
Best Local Similarity 25.3%; Pred. No. 5.6;
Matches 49; Conservative 29; Mismatches 92; Indels 24; Gaps 9;

QY 16 VLSACSTPIPKNNPOLAOI--RTQTAISLDMGKLDQAKQOL-DAASADQOPAPAYR 72
DB 57 VVPASIAIPVSVKPELITLIVSGTAYISIKELGSLSEKTVLVTAAGTGQGFAMQLS 116
QY 73 TLAVYQASEDATHTQTAQRLLFEKAI-----ELNPKD-----MOSYMDYGFYLVQMGD 121
DB 117 KKAKCHVIGTCSDEKSA--FLKSLGCDRPIKTEPVGTVLKQETPE-GVDVY-YESV 171

QY 122 SGALIFDKPSRAIGYGRVAIENMAYIYHOYEAKSPKDDYNNAKALERALISGT 181
 DB 172 GGAM--FDLAVDALATGRILVI---GFIGYQPTPLSPVKAGTLPAKLKKSASVQGF 226
 QY 182 QHDEIKKSYDKLS 195
 DB 227 FLNHVLSKYQAAMS 240

RESULT 39

US-09-134-001C-3179
 ; Sequence 3179, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3179
 ; LENGTH: 415
 ; TYPE: PRP
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3179

Query Match 7.1%; Score 74; DB 4; Length 415;
 Best Local Similarity 18.2%; Pred. No. 8.9;
 Matches 25; Conservative 24; Mismatches 54; Indels 34; Gaps 2;

QY 28 KNNPQLAQIRTOAISL-----LDMKLDQAKQOLDALSLADROFAPARYRTLAKYQASD 83
 DB 259 KEGRLRSQFYKEMLVSTGSLIEIHGDANEGVELLKQALEVDNMAHEPDLILSLDYRNEED 318
 QY 84 -----ATHQTKAQLPEKALIELNPKDMQSYMVG 113
 DB 319 YESIILQLTYVDEBDLPITFMHLAVAYGGERDKXQHFELAYPTLQTOVERLSDIYF 378
 QY 114 YLVQMGDLGALIFYDK 130
 DB 379 YLEIGQKEKAILILNQ 395

RESULT 40

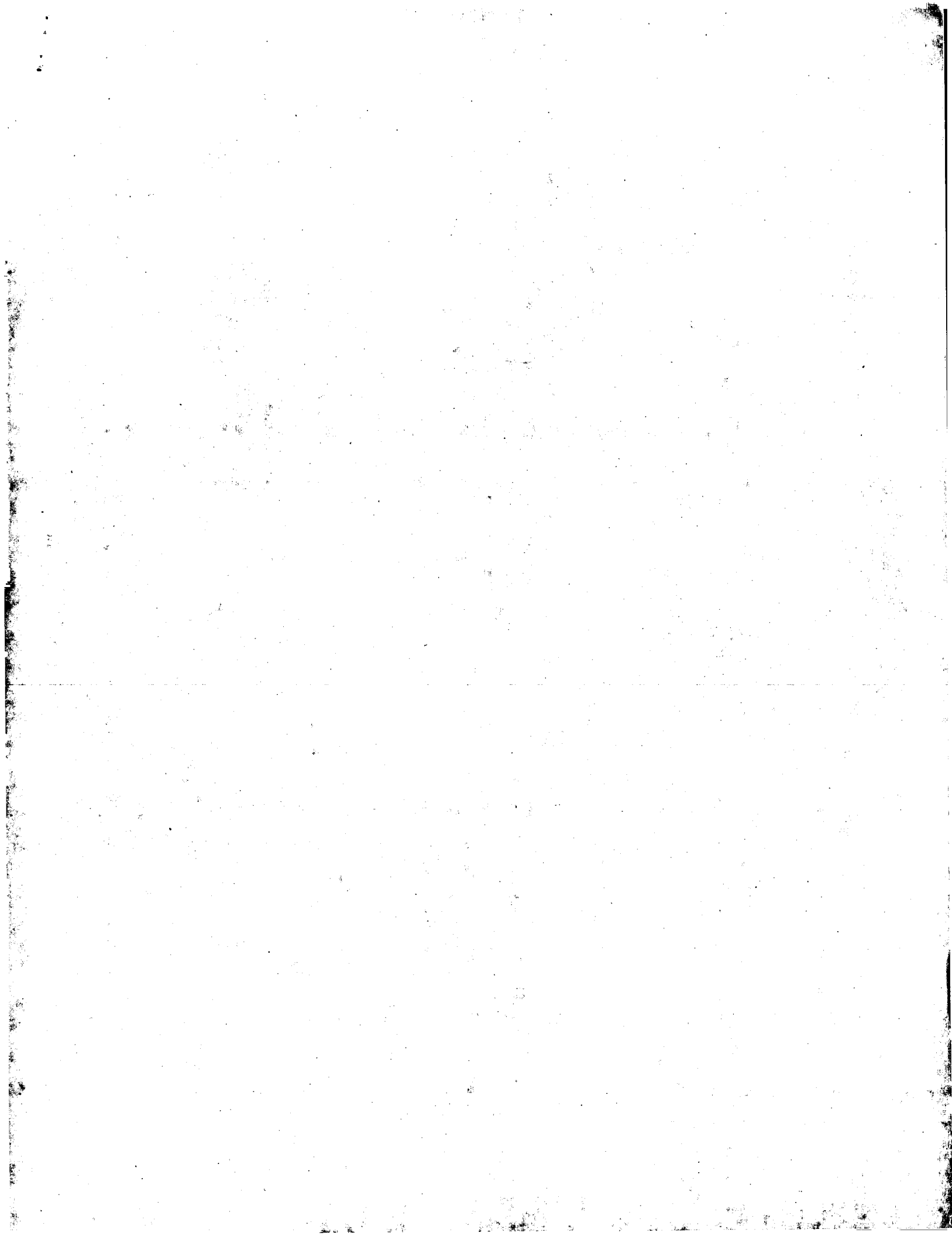
US-08-557-122A-38
 ; Sequence 38, Application US/08557122A
 ; Patent No. 5879664
 ; GENERAL INFORMATION:
 ; APPLICANT: Hjort, Carsten Mailand
 ; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
 ; NUMBER OF SEQUENCES: 38
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5879664Q No. 5879664disk of No. 5879664th America, Inc.
 ; STREET: 405 Lexington Avenue, 64th Floor
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/557,122A
 ; FILING DATE: 11-DEC-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 3980.204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 INFORMATION FOR SEQ ID NO: 38:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 638 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-557-122A-38

Query Match 7.1%; Score 74; DB 2; Length 638;
 Best Local Similarity 19.2%; Pred. No. 16;
 Matches 52; Conservative 40; Mismatches 93; Indels 86; Gaps 11;

QY 1 MKIRVMPVMMAGL--VLSAC-----OSTPIPPKNNPQLAQIRTOAISL 45
 DB 1 MKIRKMLVLLALVQLAASAGDAQEDTSPTENATEEEEDDDDLVEKENGWVL 60
 QY 46 DMGKLDQAKQOLDAL-----SADROFAPARYRTLAKYQASEDATHTQTKAQLPEKA 97
 DB 61 NDGNFDFVADKDTVLLEFYAPWCGCHKQAFAPERYEKIAGTLKNDPPIAVAK----- 112
 QY 98 IELNPKDMQSYMVGFLVQMGDLG--ALIFPKPSRAIGYGRVAIENMAYI----- 150
 DB 113 IDATSASM-----LASKFVSGYPTIKLIKQOAVDYDSSRTOEIVAKREVVSQ 163
 QY 151 -YHOYEAKSPKDDY---NNA-----KSNLE----- 174
 DB 164 DWTPPEVTLISLKNFDDVNNADIIIVEFYAPWCGCHKLAPERYEKAKEIKSPPI 223
 QY 175 -RALISGTHDEIKSYDKLSDYKLSDYK 204
 DB 224 PLAKVDATQTDIAKRFD--VSGYPTLKIFR 252

Search completed: July 8, 2003, 11:34:39
 Job time: 29 secs



RESULT 2
US-09-919-039-76
; Sequence 76, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 76
; LENGTH: 920
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 3766715CD1
US-09-919-039-76

Query Match 9.4%; Score 98.5; DB 9; Length 920;
Best Local Similarity 26.1%; Pred. No. 0.36; Matches 48; Conservative 27; Mismatches 74; Indels 35; Gaps 8;

QY 29 NNPOLAQRTOAISLDMGKLDQAKQOQDADALSDROFAPAYRTLAKYQASEDATHQ 88
DB 205 NN-LANIKR-----EQGNEBAVRLYRKALEVFPEFAAHSVLAVLQ-----QGG 249
QY 89 KAQRLEFKAIEINPKDMOSYMDYGFYLVOMGDLGSLIYFDKPSRAIGEGRVVAI 144
DB 250 KLOEALMHYKKAIRISPTFADAYSMTGNTLKMQDVQALQCY--TRAIQINPAFADAH 306
QY 145 ENMAYIY-----YHOYEAKSPTKDDYNNANSALERALISGTOHDEIKKSYDX 192
DB 307 SNLSIHDSGNIPFAISYRTALKKEDFPAYCNLAHCLO-IVCDWTIDERMKKLV 365
QY 193 LLSD 196
DB 366 IVAD 369

RESULT 3
US-10-197-666A-154
; Sequence 154, Application US/10197666A
; Publication No. US20030092037A1
; GENERAL INFORMATION:
; APPLICANT: ASAMI KASEI KABUSIKI KAISYA
; TITLE OF INVENTION: Elki phosphorylation related gene
; FILE REFERENCE: PH-1548US
; CURRENT APPLICATION NUMBER: US/10/197,666A
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: JP 2001-218204
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: JP 2001-263450
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: JP 2002-012176
; PRIOR FILING DATE: 2002-01-21
; PRIOR APPLICATION NUMBER: US 60/305,884
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/316,304
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: US 60/350,027
; PRIOR FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 154
; LENGTH: 806
; TYPE: PRT

; ORGANISM: Homo sapiens
US-10-197-666A-154

Query Match 8.9%; Score 92.5; DB 9; Length 806;
Best Local Similarity 25.5%; Pred. No. 1.2; Matches 35; Conservative 26; Mismatches 69; Indels 7; Gaps 4;

QY 28 KNPOLAQRTOAISLDMGKLDQAKQOQDADALSDROFAPAYRTLAKYQASEDATHQ 87
DB 227 KNP-IAMLYK--GLTFPHRGLLKEAIESFEKALKQKVDFTDAVKSIGQARELGNEAA 283
QY 88 TKAQRLEFKAIEINPKDMOSYMDYGFYLVOMGDLGSLIYFDKPSRAIGEGRVVAIEM 147
DB 284 TES---FOKALLNQHNVQTLQIRGMMLYHHGSLQELAKNPKKCLQLEPYNEVCQYMKGL 340
QY 148 AYIYHQ-YEAKSPTK 163
DB 341 SHVAMGQFYEGIKAQTK 357

RESULT 4
US-10-197-666A-152
; Sequence 152, Application US/10197666A
; Publication No. US20030092037A1
; GENERAL INFORMATION:
; APPLICANT: ASAMI KASEI KABUSIKI KAISYA
; TITLE OF INVENTION: Elki phosphorylation related gene
; FILE REFERENCE: PH-1548US
; CURRENT APPLICATION NUMBER: US/10/197,666A
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: JP 2001-218204
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: JP 2001-263450
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: JP 2002-012176
; PRIOR FILING DATE: 2002-01-21
; PRIOR APPLICATION NUMBER: US 60/305,884
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/316,304
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: US 60/350,027
; PRIOR FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 152
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-197-666A-152

Query Match 8.9%; Score 92.5; DB 9; Length 860;
Best Local Similarity 25.5%; Pred. No. 1.3; Matches 35; Conservative 26; Mismatches 69; Indels 7; Gaps 4;

QY 28 KNPOLAQRTOAISLDMGKLDQAKQOQDADALSDROFAPAYRTLAKYQASEDATHQ 87
DB 280 KNP-IAMLYK--GLTFPHRGLLKEAIESFEKALKQKVDFTDAVKSIGQARELGNEAA 336
QY 88 TKAQRLEFKAIEINPKDMOSYMDYGFYLVOMGDLGSLIYFDKPSRAIGEGRVVAIEM 147
DB 337 TES---FOKALLNQHNVQTLQIRGMMLYHHGSLQELAKNPKKCLQLEPYNEVCQYMKGL 393
QY 148 AYIYHQ-YEAKSPTK 163
DB 394 SHVAMGQFYEGIKAQTK 410

RESULT 5
US-10-197-666A-122
; Sequence 122, Application US/10197666A
; Publication No. US20030092037A1
; GENERAL INFORMATION:
; APPLICANT: ASAMI KASEI KABUSIKI KAISYA

;; TITLE OF INVENTION: Elki phosphorylation related gene
;; FILE REFERENCE: PH-1548US
;; CURRENT APPLICATION NUMBER: US/10/197,666A
;; CURRENT FILING DATE: 2002-11-18
;; PRIOR APPLICATION NUMBER: JP 2001-218204
;; PRIOR FILING DATE: 2001-07-18
;; PRIOR APPLICATION NUMBER: JP 2001-263450
;; PRIOR FILING DATE: 2001-08-31
;; PRIOR APPLICATION NUMBER: JP 2002-012176
;; PRIOR FILING DATE: 2002-01-21
;; PRIOR APPLICATION NUMBER: US 60/305,884
;; PRIOR FILING DATE: 2001-07-18
;; PRIOR APPLICATION NUMBER: US 60/316,304
;; PRIOR FILING DATE: 2001-09-04
;; PRIOR APPLICATION NUMBER: US 60/350,027
;; PRIOR FILING DATE: 2002-01-23
;; NUMBER OF SEQ ID NOS: 156
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 122
;; LENGTH: 716
;; TYPE: PRT
;; ORGANISM: Mus musculus
US-10-197-666A-122

Query Match 8.8%; Score 92; DB 9; Length 716;
Best Local Similarity 24.9%; Pred. No. 1.1;
Matches 48; Conservative 33; Mismatches 82; Indels 30; Gaps 8;

QY 30 NPOLAQRTOIAISLMDKGLDQA-----KQQLDA--ALSADQFAPARTLAKV 78
DB 129 NEELA-----IAVYLVSGGLYDEAIRHFTMLQEAIFAPEALAKQKVDFTDAYKSLQAV 183
QY 79 Q--ASEDATQTKQRLFEKAIELNPKDMOSYMDYGYLVQMGDLGALTYPDPSRAIG 136
DB 184 RELGNFPA-----ATESFOALLNQHNVOTQDLRGMLYHSGSLQALKNFKKCLQLEP 238
QY 137 YEGRAVAIENMAYIYHQ-YEAKSPK---DDYNNAKALERALISGTHDE--IKKS 189
DB 239 YNELCYQWKGSLSHVAMQPFYEGIKATKMLNDPLPQKXKSPYLARKVYLRYSRYLAH 298
QY 190 YDKLLSDYKLLSD 202
DB 299 LDTPLTEYNIDSD 311

RESULT 6

US-10-108-605-39
;; Sequence 39, Application US/10108605
;; Patent No. US20020160934A1
;; GENERAL INFORMATION:
;; APPLICANT: Broadue, Julie
;; APPLICANT: Starn, Lynn
;; APPLICANT: Bachmann, Jane
;; APPLICANT: Kamdar, Kim
;; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
;; FILE REFERENCE: 31133B
;; CURRENT APPLICATION NUMBER: US/10/108,605
;; CURRENT FILING DATE: 2002-03-27
;; PRIOR APPLICATION NUMBER: US 09/761,142
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/176,418
;; PRIOR FILING DATE: 2000-01-14
;; NUMBER OF SEQ ID NOS: 361
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 39
;; LENGTH: 489
;; TYPE: PRT
;; ORGANISM: Drosophila melanogaster
US-10-108-605-39

Query Match 8.8%; Score 91.5; DB 9; Length 489;
Best Local Similarity 25.0%; Pred. No. 0.76;

Matches 21; Conservative 17; Mismatches 43; Indels 3; Gaps 1;
QY 28 KNPOLAQRTOIAISLMDKGLDQAQQLDALSADQFAPARTLAKVQASEATQ 87
DB 337 KNPDDPKLYSNPACTYTLAFAFDLGLKQCDTICIKDEKFIKGIYKGLQGMQ---QQ 393
QY 88 TKAQRLFEKAIELNPKDMOSYMDY 111
DB 394 SKQAAYQKALELDPNNAEABEGY 417

RESULT 7

US-10-193-950A-2
;; Sequence 2, Application US/10193950A
;; Publication No. US20030104002A1
;; GENERAL INFORMATION:
;; APPLICANT: Sparling, P. Frederick
;; APPLICANT: Thompson, Stuart
;; TITLE OF INVENTION: Antigenic iron repressible proteins from N. meningitidis related
;; FILE REFERENCE: 381-39 CIP/CONII/CPA/CON
;; CURRENT APPLICATION NUMBER: US/10/193,950A
;; CURRENT FILING DATE: 2002-12-23
;; PRIOR APPLICATION NUMBER: 09/045,177
;; PRIOR FILING DATE: 1998-03-20
;; NUMBER OF SEQ ID NOS: 4
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 2
;; LENGTH: 1115
;; TYPE: PRT
;; ORGANISM: Neisseria Meningitidis
US-10-193-950A-2

Query Match 8.6%; Score 90; DB 9; Length 1115;
Best Local Similarity 23.2%; Pred. No. 3.3;
Matches 58; Conservative 33; Mismatches 69; Indels 90; Gaps 15;

QY 1 MKIRVMPVMWAMGLVLSAQSTPIPPKNNPOLAQRTOIAISLMDGK--LDQAQO-- 56
DB 552 MRLSTDMQTQANDGIL-----TP-----SQVNAKRALVLSKAKAADAADRIA 600
QY 57 -IDALSADQFAPARTLAKVQASEDA-----THQTKAQLFEKAI--ELNPK 103
DB 601 VLDAVYGDSS-----TL--YMSSEDAINIVKVTNDTYDHLAKNIYQNLFFQTRLOP- 651
QY 104 DMOSYMDYGYLVQMG-----DLGALTFD-----KPSRAI----- 135
DB 652 -----YLNQISFKMENDFTLLDFSGLVQAFNHVKEINPQKAFVDLAEMLAYELR 701
QY 136 -GYEGRVAIENMAYIYHQYEAKSPKTDYNNNAKALERALI--SGTHDEIKKS-- 189
DB 702 SWTEGRRLAAD-----YVEAKKAGKFEDYQKVLGQETVALIATSGTDADLIQNVG 754
QY 190 --YDKLLSDY 197
DB 755 FGHNKVLSLY 764

RESULT 8

US-10-043-487-375
;; Sequence 375, Application US/10043487
;; Publication No. US20030055220A1
;; GENERAL INFORMATION:
;; APPLICANT: HYBRIGENICS
;; APPLICANT: Pierre, LEBRAIN
;; TITLE OF INVENTION: Protein-protein interactions between Shigella flexneri polypepti
;; FILE REFERENCE: B4778A
;; CURRENT APPLICATION NUMBER: US/10/043,487
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/261,130
;; PRIOR FILING DATE: 2001-01-12
;; NUMBER OF SEQ ID NOS: 561

SOFTWARE: Patentin version 3.1
; SEQ ID NO 375
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Shigella Flexneri
US-10-043-487-375

Query Match 8.4%; Score 87.5; DB 9; Length 438;
Best Local Similarity 21.0%; Pred. No. 1.6; Indels 29; Gaps 6;
Matches 37; Conservative 36; Mismatches 74

QY 48 GGLDQAKQQLDAALSADROFAPAYRTLAKVYQASEDPATHQTAORLFEKATILNPKDMQS 107
DB 80 GPKENASQCEKVKLAKYPNNYETMKILSLVYASEQEKRDIAKHKLKVTGEYDDVEA 139
QY 108 YNDGYFYLVMGDLGALIFPDKPSRAIGYEGRVVA-----IENNAIYYH----- 153
DB 140 WLELA-QLEQTDIOGALSAYGATPRL--QEKVQADVPPEILNNVGLHFRIGNLGEAK 196
QY 154 -----QYEAASPTKOD--YNNAKSA-----LERALISGTHDEIKSYDKLSY 197
DB 197 KYFLASLDRKAEAHDEHYVAISVTTSYNLARLYEAWCEPHEAKLYKNILREH 252

RESULT 9
US-10-128-714-3166
; Sequence 3166, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3166
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3166

Query Match 8.2%; Score 86; DB 9; Length 676;
Best Local Similarity 24.3%; Pred. No. 4.2; Indels 32; Gaps 8;
Matches 44; Conservative 32; Mismatches 73

QY 28 KNNPOLAQIRTOAISLDMGKLDQAKQQLDAALSADRO--FAPAYRTLA-----K 76
DB 379 KDVDRARQIYTE-CLKLIPIHKKTFPAKIMWLKAQFDIRQMDLOARKTLGQALGMCPKDK 437
QY 77 VYQASEDPATHQ-----TKAORLFEKATILNPKDMQSYNDGYFYLVMGDLGALIFDPKPS 132
DB 438 LFRGYIDLERQLEFPRCRTLYEKQIENNPANSQSWIKYALERGLDSEARAIFE--- 494
QY 133 RAIGYEGRVVAIENM--AYIYYHQYEAASPTKDDYNNAKSALERALISGTHDEIKS 189
DB 495 --LGIDQPMIDMPELWKAVIDFEYEG-----EYDVRQLYER-LLOKTDHVKWIN 544

QY 190 Y 190
DB 545 Y 545

RESULT 10
US-10-128-714-8166
; Sequence 8166, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8166
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8166

Query Match 8.2%; Score 86; DB 9; Length 676;
Best Local Similarity 24.3%; Pred. No. 4.2;
Matches 44; Conservative 32; Mismatches 73; Indels 32; Gaps 8;

QY 28 KNNPOLAQIRTOAISLDMGKLDQAKQQLDAALSADRO--FAPAYRTLA-----K 76
DB 379 KDVDRARQIYTE-CLKLIPIHKKTFPAKIMWLKAQFDIRQMDLOARKTLGQALGMCPKDK 437
QY 77 VYQASEDPATHQ-----TKAORLFEKATILNPKDMQSYNDGYFYLVMGDLGALIFDPKPS 132
DB 438 LFRGYIDLERQLEFPRCRTLYEKQIENNPANSQSWIKYALERGLDSEARAIFE--- 494
QY 133 RAIGYEGRVVAIENM--AYIYYHQYEAASPTKDDYNNAKSALERALISGTHDEIKS 189
DB 495 --LGIDQPMIDMPELWKAVIDFEYEG-----EYDVRQLYER-LLOKTDHVKWIN 544
QY 190 Y 190
DB 545 Y 545

RESULT 11
US-09-978-244A-8
; Sequence 8, Application US/09978244A
; Publication No. US20030103992A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter S
; APPLICANT: Gorman, Jonathan D.
; APPLICANT: Candia III, Albert F.
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP MEMBRANE PROTEINS
; FILE REFERENCE: 020554-000161US
; CURRENT APPLICATION NUMBER: US/09/978,244A

;; CURRENT FILING DATE: 2001-10-15
;; PRIOR APPLICATION NUMBER: US 60/310,028
;; PRIOR FILING DATE: 2001-08-03
;; PRIOR APPLICATION NUMBER: US 09/737,246
;; PRIOR FILING DATE: 2000-12-13
;; PRIOR APPLICATION NUMBER: US 09/736,969
;; PRIOR FILING DATE: 2000-12-13
;; PRIOR APPLICATION NUMBER: US 09/736,960
;; PRIOR FILING DATE: 2000-12-13
;; PRIOR APPLICATION NUMBER: US 09/736,968
;; PRIOR FILING DATE: 2000-12-13
;; PRIOR APPLICATION NUMBER: US 60/240,545
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,508
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,503
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,539
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,543
;; PRIOR FILING DATE: 2000-10-13
;; Remaining Prior Application data removed - See file wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 106
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO: 8
;; LENGTH: 1999
;; TYPE: PRT
;; ORGANISM: Mus sp.
;; FEATURE:
;; OTHER INFORMATION: CLASP-3
US-09-978-244A-8

Query Match 8.1%; Score 84.5; DB 9; Length 1999;
Best Local Similarity 22.8%; Pred. No. 27;
Matches 44; Conservative 25; Mismatches 71; Indels 53; Gaps 10;

QY 47 MGKLDQAKQOQDALSDROPAPAYRTLAKYVQASEDA---THQTKAORLFEKAIENLP 102
DB 1636 VGLLEQQAASPSMA-GMYEAVNEVYKVLIPHEANRDAKUKLSTHGKLEAFSLIVQDG 1694
QY 103 KDM-QSMWDYGFYLVQMGDSLGLIYFDKPS-----RAIGYEGR-----VVAL-- 144
DB 1695 KRMGTFRVGFYGTGKGGDLDEQEFVYKEPAITKLAEISHRLGFGYGRFGEDVLEVIKD 1754
QY 145 -----ENMAYI-----YHGYEAKSPKTD--YNNKSLERL-----IS 179
DB 1755 SNPVDCKLDPNKAYIQITVVEPFDYEMKDRITTFDKNNY-----LRRFMYCTPPTLD 1809
QY 180 GTQDEIKSYDK 192
DB 1810 GRAHGEIHEQPKR 1822

RESULT 12
US-09-815-242-13608
; Sequence 13608, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyekind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 13608
;; LENGTH: 1179
;; TYPE: PRT
;; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13608

Query Match 8.1%; Score 84; DB 10; Length 1179;
Best Local Similarity 21.6%; Pred. No. 14;
Matches 35; Conservative 32; Mismatches 73; Indels 22; Gaps 4;

QY 22 STPIPPKNNPQLAQRTOAISLDMGKLDQAKQOQDALSDROPAPAYRTLAKYVQAS 81
DB 409 SRQSQKQADQLKTKKQALIA---KEKASQKEELETAKQYKULADYQALAK-EQEE 464
QY 82 EDATHQTKAORLFEKAIENLPKDMQSYMDYGFYLVQMGDSLGLIYFDKPSRAIGYGRV 141
DB 465 QKTSYQAOQSQQLPDLRLNLKPKQAPASLENIILRNHNSFPYAGVSVLQEKDRIGGITGAV 524

QY 142 VAIEENMAYIYHGYEAKSPKTDYNNKSLERLALISGTQH 183
DB 525 --SEHLTFDMHYQ-----TALBIALGASSQH 548

RESULT 13
US-10-106-698-6315
; Sequence 6315, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: P0005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6315
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6315

Query Match 8.0%; Score 83.5; DB 9; Length 238;
Best Local Similarity 19.2%; Pred. No. 1.8;
Matches 32; Conservative 33; Mismatches 63; Indels 39; Gaps 4;

QY 35 QIRTOAISLDMGKLDQAKQOQDALSDROPAPAYRTLAKYVQASEDATAHQTKAORLP 94
DB 1 EISHNLGVCYIYKQFKQADQDLNMLNLRHDL-TYIMGKTHLEGDL---DKALEVY 56
QY 95 EKALINPKDMQSYMDYGFYLVQMGDSLGLIYFDKPSRAIGYGRVVALENNAYIYHQ 154
DB 57 KKAVERSPENTEITLTJLGLYQLG-----IYOKA 86

OY 155 YEAKSPTRKDDYNNAKSALERALISGTOHDEIKSYDKLSPYKLLS 201
DB 87 FEHLGNALTYDPTNTYKAILAAGSMQHTCD-----FDVALTKRYVA 128

RESULT 14

US-10-153-668-194
Sequence 194, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, Akio
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAWA, Kenya
TITLE OF INVENTION: STRAT6 Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 194
LENGTH: 384
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-194

Query Match 8.0%; Score 83; DB 9; Length 384;
Best Local Similarity 24.3%; Pred. No. 3.9;

Matches 42; Conservative 25; Mismatches 90; Indels 16; Gaps 6;

OY 31 POLAQIRTOAISLDMGKLDQAKQOLDALASADROFAPAYRTIAKY----QASEDATH 86
DB 208 PESALAQOKCFALYRQAYTGNNSSQIQAMKGFEEVKKPRCAEGYALYAQALTDQOQ 267
OY 87 QTKAQLFEKAIELNPKDMOSYMDYGYLVQ-MDLSGALYFDPKPSAIGEGRV-VAI 144
DB 268 FGKADWDYKCIDLEPDNATYVHKGLQLQWKKQDLDRGL---ELISKALIEIDMKCDPFA 324
OY 145 ENNAVYIYHQ--YEAKSPTRKDDYNNAKSALERALI-----SGTOHDEIKSY 190
DB 325 ETWGTIEVQGNMEKALIDMFNKAINLAKSEMMAHLYSLCDAAHAQTEVAKKY 377

RESULT 15

US-10-153-668-196
Sequence 196, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, Akio
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAWA, Kenya
TITLE OF INVENTION: STRAT6 Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403

PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 196
LENGTH: 608
TYPE: PRT
ORGANISM: Homo sapiens
US-10-153-668-196

Query Match 8.0%; Score 83; DB 9; Length 608;
Best Local Similarity 24.3%; Pred. No. 7.3;

Matches 42; Conservative 25; Mismatches 90; Indels 16; Gaps 6;

OY 31 POLAQIRTOAISLDMGKLDQAKQOLDALASADROFAPAYRTIAKY----QASEDATH 86
DB 432 PESALAQOKCFALYRQAYTGNNSSQIQAMKGFEEVKKPRCAEGYALYAQALTDQOQ 491
OY 87 QTKAQLFEKAIELNPKDMOSYMDYGYLVQ-MDLSGALYFDPKPSAIGEGRV-VAI 144
DB 492 FGKADWDYKCIDLEPDNATYVHKGLQLQWKKQDLDRGL---ELISKALIEIDMKCDPFA 548
OY 145 ENNAVYIYHQ--YEAKSPTRKDDYNNAKSALERALI-----SGTOHDEIKSY 190
DB 549 ETWGTIEVQGNMEKALIDMFNKAINLAKSEMMAHLYSLCDAAHAQTEVAKKY 601

RESULT 16

US-09-801-368-372
Sequence 372, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Bueby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amit
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 372
LENGTH: 966
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-372

Query Match 8.0%; Score 83; DB 10; Length 966;

Best Local Similarity 22.7%; Pred. No. 14;

Matches 44; Conservative 30; Mismatches 80; Indels 40; Gaps 8;

OY 21 QSTPIPKNNPOLAQIRTOAISLDMGKLDQAKQOLDALASADROFAPAYRTIAKY 76
DB 29 QQAAYVQQLDPLDTQSTAEFTWISIASIAETLGGDBAAMAYDXTLQFNPSAKALISLAH 88

```

QY 77 WYASEATHTQTAQSLFEALIEINPKQMSYMDGYLVQKGLD-----SGALLYFD 123
Db 89 LTR-SRDMFO--RAALIEBALINDELSDWATTGHCTYLLMDLDQRAYNAVQQLTHLS 145
QY 130 KPS-----RAIGYEGRVVAIEMNAVYVYHQEAAKSPTKDQVNNAKSLERALISTGQHD 184
Db 146 NPVNPVKLMHGIG-----LLYDRYQSL-----DY--AEEFAKYLELDPHFE 184
QY 185 EIKKSYDKLSLDYK 198
Db 185 KANEIYFRLLGIYIK 198

```

RESULT 17
 US-09-995-749A-12
 Sequence 12, Application US/09995749A
 Patent No. US20020155568A1
 GENERAL INFORMATION:
 APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRA
 APPLICANT: DIJKHUIZEN, LOUBERT
 APPLICANT: RAKHOU, HAKIM
 APPLICANT: LEER, ROBERT-JAN
 TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
 FILE REFERENCE: B043388-CIP
 CURRENT APPLICATION NUMBER: US/09/995, 749A
 CURRENT FILING DATE: 2001-11-29
 PRIOR APPLICATION NUMBER: 09/604,957
 PRIOR FILING DATE: 2000-06-28
 PRIOR APPLICATION NUMBER: EPO 00201871.1
 PRIOR FILING DATE: 2000-05-25
 NUMBER OF SEQ ID NOS: 19
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 12
 LENGTH: 584
 TYPE: PRT
 ORGANISM: Leuconostoc mesenteroides
 US-09-995-749A-12

	Query Match	7.9%	Score 82.5;	DB 9;	Length 564;
	Best Local Similarity	23.1%	Pred. No. 7.8;		
	Matches	43;	Conservative	28;	Mismatches 62; Indels 53; Gaps 10
Qy	42	ISLID--MGKLDQ-AKQQLDALSADROFAPARYRLAKVYQASEDATHTQARLFKA	I	98	
Dd	79	LSIILEDWNGKDPQVYNQQGNALITMD-----YTPTS--QFGNSILTHGAN-----	I	120	
Qy	99	ELNPRDMQSNVDYGFIYLWGMDLSGALIYFDKS-----RAIGYGRVALENNAYIYYH	I	153	
Dd	121	--NRSNMNYFLDTGYTL--NGDLNKKIIVDKKNPNSGTLVNRIANSGDTRVLPNYSFVRAH	I	176	
Qy	154	QYEAKSPFTKDQDYNNAKSALERALISGTQ-----HDEIKSYDKLLSDY	I	197	
Dd	177	DYD-AQDBIR-----KAMIDHGIIKMNQDPTFPQLAQGMFEFYKDGDNPGFKKYNDY	I	229	
Qy	198	KLLSDY 203			
Dd	230	NLPSPAY 235			

RESULT 18
 US-09-736-968A-14
 : Sequence 14, Application US/09736968A
 : Patent No. US20020165283A1
 : GENERAL INFORMATION:
 : APPLICANT: Lu, Peter
 : APPLICANT: Garman, Jonathan David
 : APPLICANT: Candia III, Albert Frederick
 : APPLICANT: Abbot Velta Corporation
 : TITLE OF INVENTION: CLASP-7 Transmembrane Protein
 : FILE REFERENCE: 020054-000611US
 : CURRENT APPLICATION NUMBER: US/09/736,968A
 : CURRENT FILING DATE: 2000-12-13

```

1 PRIOR APPLICATION NUMBER: US 60/160,860
2 PRIOR FILING DATE: 1999-10-21
3 PRIOR APPLICATION NUMBER: US 60/162,458
4 PRIOR FILING DATE: 1999-10-29
5 PRIOR APPLICATION NUMBER: US 60/170,453
6 PRIOR FILING DATE: 1999-12-13
7 PRIOR APPLICATION NUMBER: US 60/176,195
8 PRIOR FILING DATE: 2000-01-14
9 PRIOR APPLICATION NUMBER: US 60/182,286
10 PRIOR FILING DATE: 2000-02-14
11 PRIOR APPLICATION NUMBER: US 09/547,276
12 PRIOR FILING DATE: 2000-04-11
13 PRIOR APPLICATION NUMBER: US 60/196,267
14 PRIOR FILING DATE: 2000-04-11
15 PRIOR APPLICATION NUMBER: US 60/196,466
16 PRIOR FILING DATE: 2000-04-11
17 PRIOR APPLICATION NUMBER: US 60/196,527
18 PRIOR FILING DATE: 2000-04-11
19 PRIOR APPLICATION NUMBER: US 60/196,528
20 PRIOR FILING DATE: 2000-04-11
21 PRIOR APPLICATION NUMBER: US 09/687,837
22 PRIOR FILING DATE: 2000-10-13
23 PRIOR APPLICATION NUMBER: US 60/240,503
24 PRIOR FILING DATE: 2000-10-13
25 PRIOR APPLICATION NUMBER: US 60/240,508
26 PRIOR FILING DATE: 2000-10-13
27 PRIOR APPLICATION NUMBER: US 60/240,539
28 PRIOR FILING DATE: 2000-10-13
29 PRIOR APPLICATION NUMBER: US 60/240,543
30 PRIOR FILING DATE: 2000-10-13
31 NUMBER OF SEQ ID NOS: 115
32 SOFTWARE: PatentIn Ver. 2.1
33 SEQ ID NO 14
34 LENGTH: 1220
35 TYPE: PRT
36 ORGANISM: Homo sapiens
37 FEATURE:
38 OTHER INFORMATION: human CLASP-3
39 US-09-736-968A-14

```

Query Match	7.9% ; Score 82.5 ; DB 9 ; Length 1220 ;
Best Local Similarity	22.6% ; Pred. No. 21 ;
Matches	44 ; Conservative 24 ; Mismatches 72 ; Indels 55 ; Gaps 10 ;
Qy	47 MGKIDQAKQOOLDAALISADROPAPARYRTLAKYQASEDA---THQTAKQRLFEKATIELN 101
Db	895 VGLLEQQAASFSMA-GWYEAVNEVYKVLIPHEANRDAKSLSTHGKQAEAFSKIVHOST 953
Qy	102 --PKDMQSYMDGYGLYLQWMDLSGALTYEDPS-----RAIGYGR-----VVAI 144
Db	954 GWERMFGTYFRVGGYGTGTFGDLDEQEFVYKEPALTTKLAELHRLBGFGEFGEDVVEVT 1013
Qy	145 -----ENMAYI-----YHQYEAAKSPTKD--YNNAKSALERAL----- 177
Db	1014 KDSNPVDKCKLDPKAKYIQLTYVEPEFDYEMKORITFYDIGNYN-----LRRFMCTPPT 1066
Qy	178 ISGQHDEIKKSYDK 192
Db	1069 LDGRAGSELHQFPR 1083

RESULT 19
US-09-736-969A-14
Sequence 14, US20020068302A1
Patent No. US20020068302A1
GENERAL INFORMATION:
APPLICANT: Lu, Peter
APPLICANT: Garman, Jonathan David
APPLICANT: Candia, III, Albert Frederick
APPLICANT: Arbor Vita Corporation
TITLE OR INVENTION: CLASP-4 Transmembrane Protein
FILE REFERENCE: 020054-000411US
CURRENT APPLICATION NUMBER: US/09/736,969A

```
;; CURRENT FILING DATE: 2000-12-13
;; PRIOR APPLICATION NUMBER: US 60/160,860
;; PRIOR FILING DATE: 1999-10-21
;; PRIOR APPLICATION NUMBER: US 60/162,498
;; PRIOR FILING DATE: 1999-10-29
;; PRIOR APPLICATION NUMBER: US 60/170,453
;; PRIOR FILING DATE: 1999-12-13
;; PRIOR APPLICATION NUMBER: US 60/176,195
;; PRIOR FILING DATE: 2000-01-14
;; PRIOR APPLICATION NUMBER: US 60/182,296
;; PRIOR FILING DATE: 2000-02-14
;; PRIOR APPLICATION NUMBER: US 09/547,276
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 60/196,267
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 60/196,460
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 60/196,527
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 60/196,528
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 09/687,837
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,503
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,508
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,539
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,543
;; PRIOR FILING DATE: 2000-10-13
;; NUMBER OF SEQ ID NOS: 153
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 14
;; LENGTH: 1220
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: human CLASP-3
US-09-736-969A-14

Query Match 7.9%; Score 82.5; DB 10; Length 1220;
Best Local Similarity 22.6%; Pred. No. 21;
Matches 44; Conservative 24; Mismatches 72; Indels 55; Gaps 10;

QY 47 MGKLDQAKQQLDALSADROFAPAYRTLAKVYQASEDA-----THQTKAQLFEKALIELN- 101
DB 895 VGLLEQQAASFSA-GMYEAVNEVYKVLPIHEANRDAKLSITHGKLQEAFSKIVHOST 953
QY 102 --PKMOSYMDYGFYLVQMGDLSCALIFYDKPS-----RAIGYGR-----VVAI- 144
DB 954 GMEHMFQYFRVGFYGTGFGDLDEQEFYKEPAITKLAESHRLEGFYGERGDEVVEVI 1013
QY 145 -----ENNAVY-----YHGYEAKSPKYD--YNNAKSALERL----- 177
DB 1014 KDSNPVDCKLDPNKAVIQITVEYPDYEMKDRITFDKNYN-----LRFMYCTPFT 1068
QY 178 ISGTQHDEIKKSYDK 192
DB 1069 LDGRAHGEHLHQFKR 1083

RESULT 20
US-09-736-960-14
; Sequence 14, Application US/09736960
; Patent No. US20020102267A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-5 Transmembrane Protein
; FILE REFERENCE: 020054-000511US
```

```
;; CURRENT APPLICATION NUMBER: US/09/736,960
;; CURRENT FILING DATE: 2001-09-20
;; PRIOR APPLICATION NUMBER: US 60/160,860
;; PRIOR FILING DATE: 1999-10-21
;; PRIOR APPLICATION NUMBER: US 60/162,498
;; PRIOR FILING DATE: 1999-10-29
;; PRIOR APPLICATION NUMBER: US 60/170,453
;; PRIOR FILING DATE: 1999-12-13
;; PRIOR APPLICATION NUMBER: US 60/176,195
;; PRIOR FILING DATE: 2000-01-14
;; PRIOR APPLICATION NUMBER: US 60/182,296
;; PRIOR FILING DATE: 2000-02-14
;; PRIOR APPLICATION NUMBER: US 09/547,276
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 60/196,267
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 60/196,460
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 60/196,527
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 60/196,528
;; PRIOR FILING DATE: 2000-04-11
;; PRIOR APPLICATION NUMBER: US 09/687,837
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,503
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,508
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,539
;; PRIOR FILING DATE: 2000-10-13
;; PRIOR APPLICATION NUMBER: US 60/240,543
;; PRIOR FILING DATE: 2000-10-13
;; NUMBER OF SEQ ID NOS: 134
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 14
;; LENGTH: 1220
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: human CLASP-3
US-09-736-960-14

Query Match 7.9%; Score 82.5; DB 10; Length 1220;
Best Local Similarity 22.6%; Pred. No. 21;
Matches 44; Conservative 24; Mismatches 72; Indels 55; Gaps 10;

QY 47 MGKLDQAKQQLDALSADROFAPAYRTLAKVYQASEDA-----THQTKAQLFEKALIELN- 101
DB 895 VGLLEQQAASFSA-GMYEAVNEVYKVLPIHEANRDAKLSITHGKLQEAFSKIVHOST 953
QY 102 --PKMOSYMDYGFYLVQMGDLSCALIFYDKPS-----RAIGYGR-----VVAI- 144
DB 954 GMEHMFQYFRVGFYGTGFGDLDEQEFYKEPAITKLAESHRLEGFYGERGDEVVEVI 1013
QY 145 -----ENNAVY-----YHGYEAKSPKYD--YNNAKSALERL----- 177
DB 1014 KDSNPVDCKLDPNKAVIQITVEYPDYEMKDRITFDKNYN-----LRFMYCTPFT 1068
QY 178 ISGTQHDEIKKSYDK 192
DB 1069 LDGRAHGEHLHQFKR 1083

RESULT 21
US-09-736-968A-107
; Sequence 107, Application US/09736968A
; Patent No. US20020169283A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-7 Transmembrane Protein
```


PRIOR APPLICATION NUMBER: US 60/170,453
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 60/176,195
PRIOR FILING DATE: 2000-01-14
PRIOR APPLICATION NUMBER: US 60/182,296
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: US 09/547,276
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,267
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,460
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,527
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,528
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 09/687,837
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,503
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,508
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,539
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,543
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 153
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 93
LENGTH: 2090
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human CLASP-3
US-09-736-969A-93

Query Match 7.9% Score 82.5; DB 10; Length 2090;
Best Local Similarity 22.6%; Pred. No. 45;
Matches 44; Conservative 24; Mismatches 72; Indels 55; Gaps 10;
QY 47 MGKLDQAKQDLADLSADROFAPAYRTLAKVYQASEDA-----THQTKAQLFEKALIELN- 101
DB 1765 VGLLEQAAAFSMA-GWEAVNEVYKVLIPHEANRDAKSLTHGKLGAEAFSKIVHQT 1823
QY 102 --PKDMQSYMDYGYFLVQMGDLSCALIFPKPS-----RAIGYEGR-----VVAI- 144
DB 1824 GWERMFGTYFRVGYGTFKFDLDEQEFYKEPAITKLAISHRLEGFYGERFGEDVVEVI 1883
QY 145 -----ENNAVY-----YHGYEAAKSPTKD--YNNAKSALERAL----- 177
DB 1884 KDSNPVDKCLDPKAKAIQTITVEPYPTIEMKDRITTFDKKNY-----LRFRMYCTPFT 1938
QY 178 ISGTQHDDEIKRSYDK 192
DB 1939 LDGRAHGEHLHGFQKR 1953

RESULT 24
US-09-736-960-90
Sequence 90, Application US/09736960
Patent No. US20020102267A1
GENERAL INFORMATION:
APPLICANT: Lu, Peter
APPLICANT: Garman, Jonathan David
APPLICANT: Candia III, Albert Frederick
APPLICANT: Arbor Vita Corporation
TITLE OF INVENTION: CLASP-5 Transmembrane Protein
FILE REFERENCE: 020054-000511US
CURRENT APPLICATION NUMBER: US/09/736,960
CURRENT FILING DATE: 2001-09-20
PRIOR APPLICATION NUMBER: US 60/160,860
PRIOR FILING DATE: 1999-10-21
PRIOR APPLICATION NUMBER: US 60/162,498

PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: US 60/170,453
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 60/176,195
PRIOR FILING DATE: 2000-01-14
PRIOR APPLICATION NUMBER: US 60/182,296
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: US 09/547,276
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,267
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,460
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,527
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,528
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 09/687,837
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,503
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,508
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,539
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,543
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 134
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 90
LENGTH: 2090
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human CLASP-3
US-09-736-960-90

Query Match 7.9% Score 82.5; DB 10; Length 2090;
Best Local Similarity 22.6%; Pred. No. 45;
Matches 44; Conservative 24; Mismatches 72; Indels 55; Gaps 10;
QY 47 MGKLDQAKQDLADLSADROFAPAYRTLAKVYQASEDA-----THQTKAQLFEKALIELN- 101
DB 1765 VGLLEQAAAFSMA-GWEAVNEVYKVLIPHEANRDAKSLTHGKLGAEAFSKIVHQT 1823
QY 102 --PKDMQSYMDYGYFLVQMGDLSCALIFPKPS-----RAIGYEGR-----VVAI- 144
DB 1824 GWERMFGTYFRVGYGTFKFDLDEQEFYKEPAITKLAISHRLEGFYGERFGEDVVEVI 1883
QY 145 -----ENNAVY-----YHGYEAAKSPTKD--YNNAKSALERAL----- 177
DB 1884 KDSNPVDKCLDPKAKAIQTITVEPYPTIEMKDRITTFDKKNY-----LRFRMYCTPFT 1938
QY 178 ISGTQHDDEIKRSYDK 192
DB 1939 LDGRAHGEHLHGFQKR 1953

RESULT 25
US-10-128-714-3384
Sequence 3384, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wengqi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshtkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714

Query Match	7.9%	Score 82	DB 9	length 689
Best Local Similarity	21.1%	Pred. NO. 11		
Matches 30	Conservative 29	Mismatches 51	Indels 32	Gaps 5

RESULT 28
US-10-027-806-16

Sequence 16, Application US/10027806
Patent No. US20020160476A1
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
FILE REFERENCE: DCOF.002A
CURRENT FILING DATE: 2001-12-21
PRIORITY FILING DATE: 1999-09-29
PRIORITY FILING DATE: EARLIER FILING DATE: 1999-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 271
TYPE: PRN
ORGANISM: Cenarchaeum symbiosum
US-10-027-806-16

Query Match 7.7%; Score 80.5; DB 9; Length 271;
Best Local Similarity 22.8%; Pred. No. 4.3;
Matches 43; Conservative 30; Mismatches 83; Indels 33; Gaps 6;

QY 28 KNNPQAOIRTOIAISLDMGKLDQAKOOLDAALSADROPAPAYRTILAKYQASEDAHQ 87
DB 37 KDDPQNRVVLQKRGALNRRYSDAITCFDILLEDDGDAFAYNNKA-IAQAEIGDT-- 93
QY 88 TKAQRLFEKALINPKDMOSYMDYGFYLVOMGDISGALYFDPKPSR----- 133
DB 94 ASALENYGRAIEASPRAPAYFNRAVLLDRIGHEHDALPDLDKATRLDRDKANPRFYKGI 153
QY 134 AIGYGRVVAIENMAYIYHGYEAASPTKDDYNNAKS---ALERALLSGTGHDEIKS 189
DB 154 VLGMGR-----HAEALSCFEVCRAHDGHADSOFHVAIEVAELG--KHAELALGE 201
QY 190 YDKLSDYK 198
DB 202 LAALPAEYR 210

RESULT 29
US-10-034-623-16
Sequence 16, Application US/10034623
Patent No. US20020198365A1
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
FILE REFERENCE: DCOF.002A
CURRENT FILING DATE: 2001-12-21
PRIORITY FILING DATE: 1999-09-29
PRIORITY FILING DATE: 1999-09-29
PRIORITY FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 271
TYPE: PRN
ORGANISM: Cenarchaeum symbiosum
US-10-034-623-16

Query Match 7.7%; Score 80.5; DB 9; Length 271;
Best Local Similarity 22.8%; Pred. No. 4.3;
Matches 43; Conservative 30; Mismatches 83; Indels 33; Gaps 6;

QY 88 TKAQRLFEKALINPKDMOSYMDYGFYLVOMGDISGALYFDPKPSR----- 133
DB 94 ASALENYGRAIEASPRAPAYFNRAVLLDRIGHEHDALPDLDKATRLDRDKANPRFYKGI 153
QY 134 AIGYGRVVAIENMAYIYHGYEAASPTKDDYNNAKS---ALERALLSGTGHDEIKS 189
DB 154 VLGMGR-----HAEALSCFEVCRAHDGHADSOFHVAIEVAELG--KHAELALGE 201
QY 190 YDKLSDYK 198
DB 202 LAALPAEYR 210

RESULT 30
US-10-027-801-16
Sequence 16, Application US/10027801
Patent No. US20030054364A1
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
FILE REFERENCE: DCOF.002A
CURRENT FILING DATE: 2001-12-21
PRIORITY FILING DATE: 1999-09-29
PRIORITY FILING DATE: EARLIER FILING DATE: 1999-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 271
TYPE: PRN
ORGANISM: Cenarchaeum symbiosum
US-10-027-801-16

Query Match 7.7%; Score 80.5; DB 9; Length 271;
Best Local Similarity 22.8%; Pred. No. 4.3;
Matches 43; Conservative 30; Mismatches 83; Indels 33; Gaps 6;

QY 28 KNNPQAOIRTOIAISLDMGKLDQAKOOLDAALSADROPAPAYRTILAKYQASEDAHQ 87
DB 37 KDDPQNRVVLQKRGALNRRYSDAITCFDILLEDDGDAFAYNNKA-IAQAEIGDT-- 93
QY 88 TKAQRLFEKALINPKDMOSYMDYGFYLVOMGDISGALYFDPKPSR----- 133
DB 94 ASALENYGRAIEASPRAPAYFNRAVLLDRIGHEHDALPDLDKATRLDRDKANPRFYKGI 153
QY 134 AIGYGRVVAIENMAYIYHGYEAASPTKDDYNNAKS---ALERALLSGTGHDEIKS 189
DB 154 VLGMGR-----HAEALSCFEVCRAHDGHADSOFHVAIEVAELG--KHAELALGE 201
QY 190 YDKLSDYK 198
DB 202 LAALPAEYR 210

RESULT 31
US-09-866-582-16
Sequence 16, Application US/09866582
Patent No. US20020127620A1
GENERAL INFORMATION:
APPLICANT: Witman, George B.
APPLICANT: Pazour, Gregory J.
APPLICANT: Rosenbaum, Joel L.
APPLICANT: Cole, Douglas G.
TITLE OF INVENTION: INTRAFACELLAR TRANSPORT
FILE REFERENCE: 07917-14501
CURRENT FILING DATE: 2001-05-24
PRIORITY FILING DATE: 2000-05-24
PRIORITY FILING DATE: 2000-05-24
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 16
LENGTH: 782
TYPE: PRT
ORGANISM: Chlamydomonas reinhardtii
US-09-866-582-16

Query Match 7.7%; Score 80.5; DB 10; Length 782;

Best Local Similarity 25.2%; Pred. No. 18;
Matches 36; Conservative 22; Mismatches 70; Indels 15; Gaps 4;

QY 28 KNNPQLAQIRTOAISLDMGKLDQAKQDADALSADROFAPAYRTLAKYQASDPATH 86
DB 432 KKEERVAKRAATNLAFLYFLEGETDQADKYSEMALKSDRYAKYVKGCVLVERGDL-- 489
QY 87 QTKAORLFEKALIELNPKDMQSYMDYGFYLVOMGDLGALTYFDKPSRAIGYEGRVAIEN 146
DB 490 -EGARSLEFNEAGIDPFCVEAIVNLGLVSRNLNLPYLAALFKLHNV--PDNVEVIHQ 546
QY 147 MAYIYVHQYEAASKPTKDDYNN 169
DB 547 IATY-----DMKGDFNGA 560

RESULT 32

US-09-815-242-13262
Sequence 13262, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:

APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2001-03-21
PRIORITY FILING DATE: 2001-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-12-22
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13262
LENGTH: 1179
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13262

Query Match 7.7%; Score 80; DB 10; Length 1179;
Best Local Similarity 21.6%; Pred. No. 36;
Matches 35; Conservative 30; Mismatches 75; Indels 22; Gaps 4;

QY 22 STPIPKNPPQALQIRTOAISLDMGKLDQAKQDADALSADROFAPAYRTLAKYQAS 81
DB 409 SRSLSQKQADQLLEKQELATRA---KEASQKQKDELFRKXQVQKLDADYDAIK-EESE 464
QY 82 EDATHQTKAORLFEKALIELNPKDMQSYMDYGFYLVOMGDLGALTYFDKPSRAIGYEGRV 141

DB 465 QKTSYQAQSQQLFDRDLNLKNKQARAQSLNLRNHSNFYAGKSVLQEKRLGILIGAV 524
QY 142 VAENMAYIYHQYEAASKPTKDDYNNAKSALRRALISGTQH 183
DB 525 --SEHLTFDYVYQ-----TALRILALGASSQH 548

RESULT 33

US-09-946-805-8
Sequence 8, Application US/09946805
Patent No. US20020116734A1
GENERAL INFORMATION:

APPLICANT: Dickman, Martin B.
TITLE OF INVENTION: PLANT DERIVED BAG HOMOLOGUES
FILE REFERENCE: 480140.469
CURRENT APPLICATION NUMBER: US/09/946,805
CURRENT FILING DATE: 2001-09-04
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 829
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-946-805-8

Query Match 7.6%; Score 79.5; DB 10; Length 829;
Best Local Similarity 24.5%; Pred. No. 25;
Matches 47; Conservative 38; Mismatches 60; Indels 47; Gaps 12;

QY 32 QLAQIRTOI--AISLDM-----GKLDQA-KQDADALSADROFAPAYRTLA-K 76
DB 380 ELGALKKKKIPFVISKDKIMDEKFEVSEKNDAGLKRQDLSLLENRQLKSDAEEK 439
QY 77 VYQASF-DATHQTKAORLFEKALIELNPKDMQSYMD-YGFYLV-----QMGDLSGA 124
DB 440 MSQLSQAEADHQELIRL-ETDVEDSRNEASIVEDYGCVFTEFGQIKCTKQETDLEHS 498
QY 125 LIYFDKPSRAIGYEGRVVAIENMAYIYHQYEAASKPTKDDYNN--KSALRRALISGTQ 182
DB 499 ML-----KEAYLLLEDLA-----RKEARKS--KDFPDSCKSVMMERCSEVYI 541
QY 183 HDEIKSYDKLL 194
DB 542 KEAVKEAKKIV 553

RESULT 34

US-10-142-515-11
Sequence 11, Application US/10142515
Publication No. US20030078399A1
GENERAL INFORMATION:

APPLICANT: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH
APPLICANT: Lloyd, Kenneth O.
APPLICANT: Yin, Beatrice W.T.
TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Uses
FILE REFERENCE: 649-A-US
CURRENT APPLICATION NUMBER: US/10/142,515
CURRENT FILING DATE: 2002-07-23
PRIORITY FILING DATE: 2002-07-23
PRIORITY FILING DATE: 2001-05-11
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 5877
TYPE: PRT
ORGANISM: Human Being
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (1) (5877)
OTHER INFORMATION: Amino acid sequence of NUC16B
US-10-142-515-11

Query Match 7.6%; Score 79.5; DB 9; Length 5877;
Best Local Similarity 21.6%; Pred. No. 36;
Matches 35; Conservative 30; Mismatches 75; Indels 22; Gaps 4;

US-09-978-244A-12

Query Match 7.6%; Score 79; DB 9; Length 2032;
Best Local Similarity 23.6%; Pred. No. 97;
Matches 37; Conservative 19; Mismatches 45; Indels 56; Gaps 8;

QY 47 MGKIDQAKQDLALSDROFAPAYRTLAKVY-----QASBD-----ATHQTKAQRLF 94
DB 1662 VGLLEQAAELFST-----GGLEYTVNEVYKLVIPILBAHRDFRLTSTHD-KLQKAF 1712
QY 95 EKALINPKDM-QSYMDYGFVLVQMDLSGALIFDPS-----PAIGYGRVVAI 144
DB 1713 DNINIKDKHMFYFRVYGFSGRFGDLDEQEFVYKSPATTKLPEISHRLEGFYGCQFGA 1772
QY 145 E-----NNAYI-----YYHOYE 156
DB 1773 EPVEVTKDTPVDKTKLDPKAKAYIQTFVBPYDEYE 1809

RESULT 38

US-08-325-278-6
Sequence 6, Application US/08325278
Publication No. US20030027283A1
GENERAL INFORMATION:
APPLICANT: Bjvreck, Lars
TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,278
FILING DATE: 26-OCT-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 450023,401
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 443 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-278-6

Query Match 7.5%; Score 78.5; DB 7; Length 443;
Best Local Similarity 27.1%; Pred. No. 13;
Matches 52; Conservative 23; Mismatches 66; Indels 51; Gaps 11;

QY 29 NNPOLAIRTOIASLMDKGLDQAKQDLAALS-ADROPAPAYRTLAKVYQASBDATHQ 87
DB 16 NNPAIQIRLR-----HENKDLKARLENVAVAGRDFKRA-EELKAKQALUED----- 62
QY 88 TKAQRLEKALINPKDMOSYMDYGFVLVQMDLSGALIFDPSRAIGYGRVVAI-- 145
DB 63 ---QR---KOLETKKELQ-QDY-----DLAKSTSDRRLKEKELEKKALELA 106
QY 146 -NNAYIYYHQ-----YEAKSFTKODYNNAKSALERALISGTQHDIEIKSY 190

DB 107 IDQASDHYRATALEKELEKKKALELAIDQASQDYNNRA-NVLEKELETTTREQEINRN- 164

QY 191 DKLLSDYKLLSD 202
DB 165 --LLGNAKLELD 174

RESULT 39

US-10-215-060-6
Sequence 6, Application US/10215060
Publication No. US20030113900A1
GENERAL INFORMATION:
APPLICANT: Alan G. Tunnicliffe
APPLICANT: David T. Welsh
APPLICANT: Bruce J. Roeder
APPLICANT: Kamaljit S. Dhalwal
APPLICANT: Camilo Colaco
TITLE OF INVENTION: METHODS OF PRESERVING PROKARYOTIC CELLS
FILE REFERENCE: 263742001701
CURRENT APPLICATION NUMBER: US/10/215,060
PRIOR FILING DATE: 2002-08-07
PRIOR APPLICATION NUMBER: US 08/985,343
PRIOR FILING DATE: 1997-12-04
PRIOR APPLICATION NUMBER: US 60/032,423
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 473
TYPE: PRT
ORGANISM: Escherichia coli
US-10-215-060-6

Query Match 7.5%; Score 78.5; DB 9; Length 473;
Best Local Similarity 22.7%; Pred. No. 15;
Matches 47; Conservative 32; Mismatches 65; Indels 63; Gaps 12;

QY 24 PIPPKNNPOLAIRTOI-AISLMDKGLDQAKQDLALSD-----RQ 66
DB 239 PLDPK---LAQKALKAVQNI-FSVERLDYSGKLPERFLAYALEKTPQHHGKIRYTO 294
QY 67 FAPAYRTLAKVYQASBDATHQAKQRLFEKALINPKDMOSYMDY-----FYLWQMG 119
DB 295 IAPTSRGDVQAYQ---DIRHQ-----LENBAGIRINK-----YQLGWTPLYILNQ-- 337
QY 120 DLSGALIFYDKP--SPAIGYE--GRVVAIENNAYIYYHOYEAKSPTK-----DDYNNNA 169
DB 338 -----HFDRKLLMKIFRYSVDGLVTPJRDGMNLVAKKEYVAADPANGVLVLSQFAG 390
QY 170 KSALEBALI-SCYQHDIEIKSYDKLLS 195
DB 391 ANELTSALIVNPFYRDVAAALDRALT 417

RESULT 40

US-10-147-026-6
Sequence 6, Application US/10147026
Publication No. US20030003538A1
GENERAL INFORMATION:
APPLICANT: Dietrich, Paul Shartzer
APPLICANT: Huang, Chiao-Chain
APPLICANT: Johnson, Carl D.
APPLICANT: Sangameswaran, Lakshmi
TITLE OF INVENTION: Neuropathic Pain Genes; Compositions
FILE REFERENCE: ROCH-006
CURRENT APPLICATION NUMBER: US/10/147,026
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: 60/155,702
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 60/189,931
PRIOR FILING DATE: 2000-04-04

Db 1 MKTISKQSAVIPPPIPSACVSGQSASSLNHQTAKARVELLSVLQNNPQLAKININDKA 60

QY 61 LSADROFPAVRYTLAKVYQASDATHQTQKQRLFEKALEINPKQMOSYMDYGYLVOMGD 120

Db 61 LQHKRYVLYVHSALAHYVQOQOQIEN--AFREYEIAVNLNHKQGDVHNNGFELCSQKK 117

QY 121 LSGALIVFDPKSPRAIGYGRVVAIENMAVITY 152

Db 118 PEOAQOQFELALNSPNYVHQADTFENIVLCAY 149

RESULT 6

C81098

Embryonal biogenesis and twitching motility protein, probable NMBJ309 [imported] - Neisseria meningitidis

C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C/Accession: C81098

R.Tetterlin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A. Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizzi, M. Science 287, 1809-1815, 2000

A/Authors: Grandl, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vettlitz, A. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. A/Reference number: AB1000; MUID:2017555; PMID:10710307

A/Accession: C81098

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-253 <TER>

A/Cross-references: GB:AE002479; GB:AE002098; NID:g7226543; PIDN:AAF41684.1; PID:g7226555

A/Experimental source: serogroup B, strain MC58

C/Genetic8:

/Gene: NMBJ309

```

Query March 11.3%; Score 117.5; DB 2; Length 253;
Best Local Similarity 23.6%; Pred. No. 0.014;
Matches 43; Conservative 42; Mismatches 78; Indels 19; Gaps 5;

Oy 4 RVKPMVMWAMGLVLSACOST--PIPPKNNPDLAQIRTOIAISLLDWMGLDQAKQOOLBAAL 61
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 RISLLLVLTALG---ACSTSYRPSRAEKANOVSNIKTOLAEHRYRGQDYKQATASIEDAL 63

Oy 62 SADRQAPAYRYTLAKVYOASEDATHQTYAQRLEFEKATLELNKDMQSYVDYGFYLV-QMGD 120
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 KSDPKNELAMLVRAEITYGLK---VNDKAGQSFPGQALSIRKDSMEINNNTYMFICGLNR 120

Oy 121 LSGALITYDKRSRAIGYGRVALENMAYIYYHQYEAASKPTKDDYNNAKSALERALLSG 180
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 PAESMAYFDKALADPTPTPYIANLN-----KGI CSAKQGFGLAEAYLKRSLAAQ 171

Oy 181 TQ 182

Db 172 PQ 173

RESULT 7
G61843
probable_lipoprotein NMA1523 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C/species: Neisseria meningitidis
C/date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C/Accession: G61843
R.Pattnilil, U.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
J.; Holtroyd, S.; Jogley, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A/Reference number: A01775, MUID:20222556, PMID:10761919
A/Accession: G61843
A/Status: preliminary
A/molecule type: DNA
A/Residues: 1-253 <PAR>
A/Cross-references: GB:AL162756, GB:AL157959, NID:g7380091, PIDN:CAB84751.1, PID:g738016
A/Experimental source: serogroup A, strain Z2491
C/Genetics:
A/Gene: NMA1523

```

	Query Match	11.3%;	Score 117.5;	DB 2;	Length 253;	
	Best Local Similarity	23.6%;	Pred. No. 0.014;			
	Matches	43;	Conservative	42;	Mismatches	78; Indels 19; Gaps 5;
OY	4 RVKPMVMANGLVISACOST--PIPKNNPOLAOIFTOIALSLDMGKLDQAKOOLDAAL	61				
Db	8 RISLLVLALG---ACSTSYRSPRAEKANOVSNIKTOLAMEYMGODPYRATASIEDAL	63				
OY	62 SADROPAPVRYTLAKVVOASEDAVHQTGAORLFEKATELNFKMQDSVMDYGTVLV-OWGD	120				
Db	64 KSDPRKELAMIVRAEITFOYLK---VNDKAQGSFPOALSIRKDSAEINNNGWFLCGRLNR	120				
OY	121 LSGALIYPDKESRAIGYEGRVALEINMAYIHYOEAAKSPTKDDYNNAKSALEBALISG	180				
Db	121 PAESMAAFEDKALADPTFTPYIANLIN-----KGICSAKOGCFGIAEAVLYKRSLAAG	171				
OY	181 TQ 182					
Db	172 PQ 173					

```

RESULT 8
T31673
N:acetylglucosaminyltransferases (EC 2.4.1.-), chain p110 - rat
N:Alternate names: O-GlcNAc transferase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
C:Accession: T31673
R:Kreppel, L.K.; Blomberg, M.A.; Hart, G.W.
J: Biol. Chem. 272, 9308-9315, 1997
A:Title: Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characteri
A:Reference number: 221056; MUID:97238869; PMID:9083067
A:Accession: T31673
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-1036 <RFE>
A:Cross-references: EMBL:U76557; NID:g1931578; PID:g1931579; PIDN:AAC53121.1
A:Experimental source: strain Sprague Dawley; liver
C:Genetics:
A:Gene: OGT
C:Keywords: glycosyltransferase; hexosyltransferase; tandem repeat

Query Match          11.3%; Score 117.5; DB 2; Length 1036;
Best Local Similarity 25.5%; Pred. No. 0.079;
Matches 48; Conservative 33; Mismatches 68; Indels 39; Gaps 7;

QY 10 VMANGVLKSCOSTPIPPKKNPOLAQIRIOAISLDMKGLDQAKQQLDAALSADROFAP 69
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 123 LVAAQMEGAIVQAVYSALQYNPDLYCVRSDLGMLTKALGRLEBAKACYCLKAIETQPNFAV 182

QY 70 AYRLTAAVYQASD---ATHQTKAQRLEKAEIENPKDQSYNDYGFYLVQWDDLSGALI 126
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 183 AMSNLGCVFPAQGIWLAHH-----FEKAVLDPDPFLDAYINLGNVLKE-----AR 230

QY 127 YFDKPSRAIGYEGRVVAIR-----NMAYIYVHQ-----YEA-----KSPTKDDY 166
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 231 F----DVAVAVYIRALSLSPNHAHVHGNLACVYEGQLDLAIDTTRRAIEIQPHRPDAV 286

QY 167 NNAKSALE 174
    |:::|
Db 287 CNLNAALK 294

RESULT 9
H97323
TFR-repeat-containing protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97323
R:Noelling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J: Bacteriol. 183, 4823-4838, 2001

```

```

Query Match      11.0%; Score 115; DB 2; Length 237;
Best Local Similarity 23.5%; Pred. No. 0.02;
Matches      50; Conservative 40; Mismatches 63; Indels 60; Gaps 11;

Qy      10 VMAMGLV-----SACQSTPIPPKNNPQLAQIRQTQIAISLSLMGKLDQAKQ 56
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      4 VFGLGLIILALAGCVTVETETAGNATQSDP-----TEMAEKRIALGGLYENGSMITKAREN 57
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      57 LDALSLDROFAPY-----RTIAKVYQASEDAHQTAORLFEKAIETLNPKDMQSYWDYG 112
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      58 LERKL-----QAPPYRSQLSMAHYTA-----VGNDSARKKRYRALSSEHRNGVLANFG 110
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      113 FYLVQMDDLSCGALFYFKPSPAIGYEGSRVAIEN-MAIYYHQYE-----AASPTQDDY 166
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      111 TFLCKQGEYDTPADQYFRR-----AVEQPPYYLLISASYENAGLCALAKGKT-- 156
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      167 NNAKSALERALISGTHGEIKKSYDKLISDYKL 199
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      157 -NAREYFRRL-----DHP-----NRLISIIQL 179
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 12
 S45825
 cell division control protein CDC27 - Yeast (Saccharomyces cerevisiae)
 N/Alternate names: protein YBL0718; protein YBL084c
 C/Species: Saccharomyces cerevisiae
 C/Date: 26-Aug-1994 #sequence revision 09-Sep-1994 #text_change 11-Jan-2000
 C/Accession: S45825; S45422; S28742; S59219
 R/Domdey, H.; Gassenhuber, H.; Obermaler, B.; Piravandi, E.
 submitted to the Protein Sequence Database, August 1994
 A/Reference number: S45816
 A/Accession: S45825
 A/Molecule type: DNA
 A/Residues: 1-758 <DOM>
 A/Cross-references: EMBL:Z55845; NID:G536135; PIDN:CA04905.1; PID:G536136; MIPS:YBL084c
 A/Experimental source: EMBL Data Library, strain S286C
 A/Experimental source: EMBL Data Library, May 1994
 submitted to the EMBL Data Library, May 1994
 A/Description: Sequence analysis of a 78,6 Kb segment of the left end of Saccharomyces cerevisiae chromosome 12
 A/Reference number: S45387
 A/Accession: S45422
 A/Molecule type: DNA
 A/Residues: 141-758 <OBE>
 A/Cross-references: EMBL:X79489; NID:G496661; PIDN:CA056022.1; PID:G496695
 A/Experimental source: EMBL Data Library, strain S286C
 R/Skorzki, R.S.; Michaud, W.A.; Wootton, J.C.; Boguski, M.S.; Connolly, C.; Hieter, P.
 Cold Spring Harb. Symp. Quant. Biol. 56, 663-673, 1991
 A/Title: YPR proteins as essential components of the yeast cell cycle.
 A/Reference number: S28742; MUID:92306611; PMID:1819514
 A/Accession: S28742

A>Status: not compared with conceptual translation
A.Molecule type: DNA
A.Residues: 607-619 <SIR>
R.Obertmiller, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
Vest 11, 1103-1112, 1995
A>Title: Sequence analysis of a 78.6 kb segment of the left end of *Saccharomyces cerevisiae*
A.Reference number: S59184; MUID:96076635; PMID:1502586
A.Accession: S59219
A>Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 442-691 <OBW>
A.Cross-references: EMBL:X79489
C.Genetics:
A.Gene: SGD: CDC27; SNI1
A.Cross-references: SGD:S0000180; MIPS:YBL084C
A.Map position: 2L
C.Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc
C.Keywords: nucleus
F.472-505/Domain: tetratricopeptide repeat homology <TT1>
F.506-539/Domain: tetratricopeptide repeat homology <TT2>
F.540-573/Domain: tetratricopeptide repeat homology <TT3>
F.574-607/Domain: tetratricopeptide repeat homology <TT4>
F.608-641/Domain: tetratricopeptide repeat homology <TT5>
F.642-675/Domain: tetratricopeptide repeat homology <TT6>
F.676-709/Domain: tetratricopeptide repeat homology <TT7>
F.710-743/Domain: tetratricopeptide repeat homology <TT8>

Query Match 10.8%; Score 113; DB 2; Length 758;
Best local Similarity 23.8%; Pred. No. 0.13;
Matches 46; Conservative 43; Mismatches 66; Indels 38; Gaps 9;

OY 38 TQAIISLD-----MGKLDQAKQOLDAALSA-----DRQAPAYRTAKYQ 79
Db 526 SNLNGMLDTPNKRKRETCIGNLSLQKDHDAIKAEKATQDLPNPAAY-TIQGHN 584
OY 80 ASEDATHTQKARLEKALIELNPKQMOSYMDGYLVQMGDISGALIYFDKPSALIGYEG 139
Db 585 SSNDSDSAKT--CYRKALACDPQHNNAYVIGTSGAMKLGQEBAALYFEK-ARSINPVN 641
OY 140 RVV-----ATENMAY-----IYHGYEAKSPTKQDYNNAKALERALISGQHDEIKKS 189
Db 642 VLLICCCGSLKELGKYEKALQYELACHLOPTS---SLSKYKQGLLYSMTRYVVALQT 698
OY 190 YDKLSDYKLSLD 202
Db 699 FBEELV---KLVPD 708

RESULT 13
B69196
Conserved hypothetical protein MTH72 - *Methanobacterium thermoautotrophicum* (strain Delta C)
C.Species: *Methanobacterium thermoautotrophicum*
C.Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jan-2000
C.Accession: B69196
R.Smith, D.R.; Doucette-Stamm, L.A.; DeLonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Kl, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A>Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: funcu
A.Reference number: A69000; MUID:98037514; PMID:9371463
A.Accession: B69196
A.Status: preliminary; nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-403 <MTH>
A.Cross-references: GB:AE000798; GB:AE000666; NID:92621094; PIDN:AA894576.1; PID:9262110
A.Experimental source: strain Delta H
C.Genetics:
A.Gene: MTH2
C.Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc
F.16-49/Domain: tetratricopeptide repeat homology <TT01>
F.50-83/Domain: tetratricopeptide repeat homology <TT02>
F.84-117/Domain: tetratricopeptide repeat homology <TT03>
F.118-151/Domain: tetratricopeptide repeat homology <TT04>

F.152-185/Domain: tetratricopeptide repeat homology <TT05>
F.186-219/Domain: tetratricopeptide repeat homology <TT06>
F.220-253/Domain: tetratricopeptide repeat homology <TT07>
F.254-287/Domain: tetratricopeptide repeat homology <TT08>
F.288-321/Domain: tetratricopeptide repeat homology <TT09>
F.322-355/Domain: tetratricopeptide repeat homology <TT10>
F.356-389/Domain: tetratricopeptide repeat homology <TT11>

Query Match 10.7%; Score 111.5; DB 2; Length 403;
Best local Similarity 24.9%; Pred. No. 0.078;
Matches 46; Conservative 34; Mismatches 60; Indels 45; Gaps 7;

OY 43 SLTD-MGKLDQAKQOLDAALSDRQAPAYRTAKYQASEDATHTQKARLEKALIELN 101
Db 228 ALDTTGKPEKALICEKALEINQKAKMNNKVVL---BELKRYDEALCEKALEIN 284
OY 102 PKDMOSYMDGYLVQMGDISGALIYFDK-----PSALIGYEGVAIENMAYIYHGYE 156
Db 285 LENDETANKGVLLRKLGKEALECEKALEINPEPADAMEWKGIILEDV----- 335
OY 157 AAKSPTQDYNNAKALERALISGTQ-----HDEIKSYDKLSDYKL 199
Db 336 --KRP-----EALKCYEKALKLNPDQKTLWYQKTLQKLGKQAKKSYKAL---KI 385
OY 200 LSDYK 204
Db 386 DPEYK 390

RESULT 14
H83171
Type 4 fibroblast biogenesis protein p11f PA3805 [imported] - *Pseudomonas aeruginosa* (stra
C.Species: *Pseudomonas aeruginosa*
C.Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C.Accession: H83171
R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.V.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A.Reference number: A82950; MUID:20437337; PMID:10984043
A.Accession: H83171
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-252 <STO>
A.Cross-references: GB:AE004798; GB:AE004091; NID:9949963; PIDN:AA607192.1; GSPDB:GN001
A.Experimental source: strain PA01
C.Genetics:
A.Gene: p11f; PA3805

Query Match 10.6%; Score 111; DB 2; Length 252;
Best local Similarity 24.4%; Pred. No. 0.048;
Matches 53; Conservative 29; Mismatches 87; Indels 48; Gaps 7;

OY 1 MKIRKPMVWAMGL---VLSACSTPIPKNNPOLAQIRQIAISLDMGKLDQAKQOL 57
Db 1 MTVAAALVFLAVGLTGTVISGDDNPKTKDGRBADAYQLGLGYVQRENTQAKYPL 60
OY 58 DAALSADRFAPAYRTAKYQAS-----EDAT 85
Db 61 RKALEIDPSSADAAALAVVQTEMPEPLADEBYRKALASDRNARVANNYGCFLYEQR 120
OY 86 HQTKAQGLFEKALF-LMPKQMOSYMDGYLVQMGDISGALIYFDKPSALIGYEGVAI 144
Db 121 YEBAVQRLEASQDTLTPERSRVEENGLVSLQKKKPPQAKYEYKSLR-LNRNOPSVAL 179
OY 145 ENMAYIYHGYEAKSPTKQDYN-----NAKSAL 173
Db 180 E-MADLLYKEREYV--PARQYVDLFAQGGCGNNASL 213

RESULT 15
D72213

C/Accession: AC2654
 R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuyavian, T.; Levy, R.; Li, M.; McClell
 , Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Author: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AC2654
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-651 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AA1649.1; PID:gt1738990; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu0633
 A:Map position: circular chromosome

Query Match 9.9%; Score 103; DB 2; Length 651;
 Best Local Similarity 21.2%; Pred. No. 0.75;
 Matches 46; Conservative 40; Mismatches 57; Indels 74; Gaps 10;

QY 27 PKNNPQALQIRTOIALSLDMGKLDQAKQDLAALSAD-----ROF-----APAYRT 73
 DB 387 PKDSPMHRISLQGLTLAQTKVESRKHLLSLSDPKDIRSYLAVGSVLSADKDYKA 446
 QY 74 LAKVYQ-----ASEDATHTQAKRLFEKALINPKDMQSYND 110
 DB 447 MAENYKAVEIGAVGQKSDMSVFQRIAYERLKQWDAEPFKALEINPEQPVLVNY 506
 QY 111 YGFYLVQMGDLSCALTYFDKP-----SRAL-----GYEGRVAIENMAYTYHQYAAKS 160
 DB 507 LGYSWVDKG-----INLDEGMKMSIRAVALELPNDY-----IVDSLGAHYR----- 548
 QY 161 PTKDDYNNAKSALEALISGTHDEIKKSYDKLLSDY 197
 DB 549 --LGDFDQSVTELEIRAI-----ELKAG-DPTINDH 575

RESULT 23

G97435
 hypothetical protein AGR_C_1124 [imported] - *Agrobacterium tumefaciens* (strain C58, Cere
 C/Species: *Agrobacterium tumefaciens*
 C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C/Accession: G97435
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*
 A:Reference number: A97359; PMID:11743194
 A:Accession: G97435
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-659 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK6440.1; PID:g15155580; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_1124
 A:Map position: circular chromosome

Query Match 9.9%; Score 103; DB 2; Length 659;
 Best Local Similarity 21.2%; Pred. No. 0.76;
 Matches 46; Conservative 40; Mismatches 57; Indels 74; Gaps 10;

QY 27 PKNNPQALQIRTOIALSLDMGKLDQAKQDLAALSAD-----ROF-----APAYRT 73
 DB 395 PKDSPMHRISLQGLTLAQTKVESRKHLLSLSDPKDIRSYLAVGSVLSADKDYKA 454
 QY 74 LAKVYQ-----ASEDATHTQAKRLFEKALINPKDMQSYND 110
 DB 455 MAENYKAVEIGAVGQKSDMSVFQRIAYERLKQWDAEPFKALEINPEQPVLVNY 514
 QY 111 YGFYLVQMGDLSCALTYFDKP-----SRAL-----GYEGRVAIENMAYTYHQYAAKS 160

DB 515 LGISWVKG-----INLDEGMKMSIRAVALELPNDY-----IVDSLGAHYR----- 556
 QY 161 PTKDDYNNAKSALEALISGTHDEIKKSYDKLLSDY 197
 DB 557 --LGDFDQSVTELEIRAI-----ELKAG-DPTINDH 583

RESULT 24

T31808
 hypothetical protein M03F8.3 - *Caenorhabditis elegans*
 C/Species: *Caenorhabditis elegans*
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 15-Sep-2000
 C/Accession: T31808
 R:Davidson, S.; Wohlmann, P.
 submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of C. elegans cosmid M03F8.
 A:Reference number: Z21089
 A:Accession: T31808
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-761 <DAV>
 A:Cross-references: EMBL:AF016441; PIDN:AA65909.1; GSPDB:GN00023; CESP:M03F8.3
 A:Experimental source: strain Bristol N2; clone M03F8
 C:Genetics:
 A:Gene: CESP:M03F8.3
 A:Map position: 5
 A:Insertions: 43/3; 388/3; 459/3; 632/2
 C/Superfamily: Schizosaccharomyces pombe hypothetical protein SPEC31F10.11c

Query Match 9.9%; Score 103; DB 2; Length 761;
 Best Local Similarity 19.8%; Pred. No. 0.91;
 Matches 37; Conservative 44; Mismatches 68; Indels 38; Gaps 6;

QY 47 MGLDQAKQDLAALSADRPAPAYRTIATKYQASDATHTQAKRLFEKALINPKDMQ 106
 DB 116 IGEIOPARSVERALDVHRISISIMLOYAEMEMCKQINH--ARNVFDRAITTPRAMQ 172
 QY 107 SYMDYGFYLVQMGDLSCALTYFDK-----PSRA-----IGYEGHVAIENMAYTY----- 151
 DB 173 FMLKSYNEEVIENTPGKQRIEFKRIEWEPEQAMQTYINELAKKEIDRRARSYQRTLH 232
 QY 152 -----HYQYEAASPTKDY--NNAKSALERA-----LISGTHDEIKSYD 191
 DB 223 VHGINVQWIKYAKKEENGYIGNARAYEKAMEYGEEDINETVLVAFALFEERQKHE 292
 QY 192 KLLSDYK 198
 DB 293 RARGIFK 299

RESULT 25

AC2388
 hypothetical protein all4659 [imported] - *Nostoc* sp. (strain PCC 7120)
 C/Species: *Nostoc* sp.
 A/Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C/Accession: AC2388
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matsumoto, A.; Iriyuchi,
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AC2388
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-236 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA076356.1; PID:gt17133736; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all4659

Query Match 9.8%; Score 102.5; DB 2; Length 236;

[illegible]

C.Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C.Accession: AF1263

R.Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgat, O.; Entian, K.D.; Fehrl, H.; Jones, L.M.; Karsch, U.

Science 294, 849-852, 2001

A.Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurupkat, G.; Madueno, E.; Maitournam, A.; Mak, O.; C.; Schlueter, T.; Sinoes, N.; Tlerez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.

File: Comparative genomics of *Listeria species*.

A.Reference number: AB1077; WUID:21537279; PMID:11679669

A.Accession: AF1263

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-212 <GLA>

A.Cross-references: GB:NC_003210; PIDN:CAC99588.1; PID:g16410939; GSPDB:GN00177

A.Experimental source: strain EGD-e

C.Genetics:

A.Gene: lmo1510

Query Match

Best Local Similarity 23.8%; Score 101.5; DB 2; Length 212;

Matches 41; Conservative 24; Mismatches 66; Indels 41; Gaps 4;

44 LIDMGKLDQAQKQDLAALSDRQFAPAYRTLAKY-----QASEDTHQ----- 87

34 LLSMDFFERAELEPFKALBLDTPPAAYYSIGMLYELERYQEAADSFQNAKQMGEND 93

88 -----TKAQLFEKATELNPCKMQSYVDYGFYLVQMGDLSGALYFDKPS 132

94 LFFMLGMSVQVMBELTLNAPYLRLSRVELNPEDGALFQYIGVLAKSG-----FYED--- 144

133 RAIGYGRVVALENMAYIYHYOEAASPTKDDYNNKASLALRALISCTQHD 184

145 -AINMLERVLTVKPEDPDALVYNIGAAVLAWGQGIYLAQYFERRALATGASHR 195

RESULT 28

F64399

hypothetical protein MJ0798 - Methanococcus jannaschii

C.Species: Methanococcus jannaschii

C.Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C.Accession: F64399

R.Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Ison, J.D.; Sadov, P.W.; Hama, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A.Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.

A.File: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A.Reference number: A64300; WUID:96337999; PMID:8688087

A.Accession: F64399

A.Status: preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-334 <BU>

A.Cross-references: GB:U67524; GB:L77117; NID:92826321; PIDN:AAB98793.1; PID:g1499620; T

C.Genetics:

A.Map position: FOR721781-722785

C.Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat ho

F102-113/Domain: tetratricopeptide repeat homology #status atypical <TII>

F135-166/Domain: tetratricopeptide repeat homology <TII>

F169-201/Domain: tetratricopeptide repeat homology #status atypical <TT3>

F202-235/Domain: tetratricopeptide repeat homology <TT4>

F236-269/Domain: tetratricopeptide repeat homology <TT5>

F272-306/Domain: tetratricopeptide repeat homology <TT6>

Query Match

Best Local Similarity 23.8%; Score 101.5; DB 2; Length 334;

Matches 40; Conservative 29; Mismatches 68; Indels 31; Gaps 5;

43 SLIDMGKLDQAQKQDLAALSDRQFAPAYRTLAKVYQASDATHQTQAQLFEKALLENP 102

144 ALTKYKQDLAKYFKAKSEKQNNYKALFGLSKSYLMSDNKNSIK--YFEKLEINLP 200

103 KDMQSYMDYGFYLVQMGDLSGALYFDKPSRAIGYGRVVAIE-----NMAYIYHYO 154

DB 201 NDVBA-LFY-----LGEIYEDCEKALNFEKALFKPDDIDLILKAFYFKL 249

QY 155 YEAKSPTKDDYNNAKSALERALISGTQHDEIKKSYDKLSDYKLSLD 202

DB 250 KK-----YKHKYFEKALKLNPNVELEQIYVSMRITIIYIGE 288

RESULT 29

F83807
hypothetical protein BH1262 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83807
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F83807
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-214 <STO>
A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BA04981.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1262

Query Match 9.6%; Score 100; DB 2; Length 214;
Best Local Similarity 40.3%; Pred. No. 0.33;
Matches 27; Conservative 9; Mismatches 27; Indels 4; Gaps 2;

QY 89 KAORLEPKALELPMKQMSYMDYGFYLVQMGDLSGLIYDKRSRRLG-YEGVVALEMM 147

DB 8 EAAKAFMAALEANPSPDPIGFVNGNLGWLGVLDKALIFPDK--AIGQEDCAPAYYGA 64

QY 148 AVIYHQ 154

DB 65 GTIYKQ 71

RESULT 30

AF1917
hypothetical protein al10889 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AF1917
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF1917
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-605 <KIR>
A:Cross-references: GB:BA000019; PIDN:BAH72846.1; PID:g17130234; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al10889

Query Match 9.5%; Score 99.5; DB 2; Length 605;
Best Local Similarity 22.2%; Pred. No. 1.3;
Matches 45; Conservative 33; Mismatches 88; Indels 37; Gaps 7;

QY 12 AMGLVLASQSTPIPKN-NPOLAQRTQIAISLDMGKLDQAKQDLAALSADROFAPA 70

DB 124 AIAVYISTIEPPLASINIEDIANAYYRGUNNDQGVQEAIIIDLOALQMHYFPA 183

QY 71 YRTIAKY-----QASEDATHQTQAQRLFEKALINFEKQMSYMDYGFYLVQMGDLSG 123

DB 184 YSIRGNYYKLGKGYRQAIAD-----HERAVQIDENLBAVYONGNNAHYALGAYOK 233

QY 124 ALIYFDK-----PSAIGYEGRVVAIENMAIYYHQYPAASPTKO-----DYNNAKSALB 174

DB 234 AIADYNRTLEINPHOVGAAYNR-----GLISFYHEYOQAFADFNQVINFNPDQAQAYQ 288

QY 175 RALISGTQHDEIKKSYDKLSDY 197

DB 289 RGLIY-----EAMQDYQSALADY 306

RESULT 31

MM7M1
myosin heavy chain D [similarity] - Caenorhabditis elegans
N:Alternate names: myosin heavy chain I
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Caenorhabditis elegans
C:Date: 28-Feb-1986 #sequence_revision 19-May-2000 #text_change 19-Apr-2002
C:Accession: T21193; T23973; S02772; A02993
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19388
A:Accession: T21193
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1938 <MTL>
A:Cross-references: EMBL:Z71261; PIDN:CAA95806.1; GSPDB:GN00019; CESP:R06C7.10
A:Experimental source: Clone F21C3
R:Gardner, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19825
A:Accession: T23973
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1938 <MT2>
A:Cross-references: EMBL:Z71266; PIDN:CAA95848.1; GSPDB:GN00019; CESP:R06C7.10
A:Experimental source: Clone R06C7
R:Dibb, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.
J. Mol. Biol. 205, 603-613, 1989
A:Title: Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain gen
A:Reference number: S02771; MUID:89178677; PMID:2926820
A:Accession: S02772
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-376 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DIB>
A:Cross-references: EMBL:X08065; NID:96785; PIDN:CAA30854.1; PID:g6786
R:Karn, J.; Brenner, S.; Barnett, L.
Proc. Natl. Acad. Sci. U.S.A. 86, 4253-4257, 1989
A:Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy ch
A:Reference number: A93958; MUID:83273600; PMID:6576334
A:Accession: A02993
A:Molecule type: DNA
A:Residues: 24-93, 'E', 95-97, 'R', 99-376, 'V', 378-388, 'GDV', 392-407, 'N', 409-473, 'G', 475-576
C:Genetics:
A:Gene: myo-1; CESP:R06C7.10
A:Map position: 1
A:Introns: 23/3; 114/3; 229/1; 264/1; 320/1; 857/3; 1745/3; 1814/1; 1892/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle co
F:87-773/Domain: myosin motor domain homology <MMOT>
F:177-184/Region: nucleotide-binding motif A (P-loop)
F:660-682/Region: actin binding #status predicted
F:764-778/Region: actin binding #status predicted
F:846-1938/Domain: coiled coil #status predicted <COI>
F:846-1160/Region: S2
F:1161-1938/Region: light meromyosin
F:128/Modified site: N6, N6, N6-trimethyllysine (Lys) #status predicted
F:183/Binding site: ATP (Lys) #status predicted
F:700, 710/Active site: Cys #status predicted

Query Match 9.5%; Score 99.5; DB 1; Length 1938;
Best Local Similarity 27.7%; Pred. No. 5.8;
Matches 49; Conservative 34; Mismatches 65; Indels 29; Gaps 9;

QY 29 NNPOLAQRTQIAISL-IDMGKLDQAKQDLAALS-ADROFAPAYRTIAKYQASEDATH 86

Db 1726 NNAALSAKKRKEVNEVOJAIARNEIDELYNELKASEERAKKAADADRLAEVYQOEQEHAVH 1785
 Qy 87 QTKORLEFKAIELNPKMOSYMDYGFYLVQMGDSLGLIYDQKSPRAIGYGRVALEN 146
 Db 1786 VDR-OR---KSLMIGLAKELQAKID-----DAERMIIGCAVALA-KYEDRRVSLR- 1830
 Qy 147 MAYIYHO---YEAAKSPTKDDYNNAKSALERALLSGTQHDRIKKSVDKLSDYKL 200
 Db 1831 -AEHSEGRRHQESIKGYTKGR-----RRARELQGVVEDKKAFLRLQENVEKL 1878

RESULT 32
 TPR-repeat-containing protein [imported] - Clostridium acetobutylicum
 A:Accession: A96951
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: A96951
 R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO
 A:Reference number: A96900; PMID:21359325; PMID:21359325
 A:Accession: A96951
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-376 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK78396.1; PID:G15023269; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC0416

Query Match 9.5%; Score 99; DB 2; Length 376;
 Best Local Similarity 22.6%; Pred. No. 0.82;
 Matches 45; Conservative 27; Mismatches 85; Indels 42; Gaps 6;

Qy 28 KNNPOLAQRTOIAISLDMGKLDQAKQOLDALSDROFAPAYRTLAKVQASEDATHQ 87
 Db 105 KNNKNDKTKALVYKGISLBRAGEYLEAIECYDILLDKMKMDAYSKARV---QEIGKH 161
 Qy 88 TKQAORLEFKAIELNPKMOSYMDYGFYLVQMGDSLGLIYDQKSPRAIGYGRVALEN 142
 Db 162 KEAISTYIAINLKKDDDEPYLRISIVRIGFYDKALYIEKLEAVPERVIGYVDKAL 221
 Qy 143 AIENNAVYIYHQYEAKSPKODYNNAKSALERALLSGT-----QHDE 185
 Db 222 IL-----ISTERYNDALEWINRAIKMENRDEPIYITIGNLYRKIKQYDE 265
 Qy 186 IKKSYDKLSS-DYKLLSDY 203
 Db 266 AIKMYDEAIKYDKSFFKSY 284

RESULT 33
 G11299
 conserved hypothetical protein TP0648 - syphilis spirochete
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
 C:Accession: G11299
 R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
 reon, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McD
 chey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 Science 281, 375-388, 1998
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
 A:Reference number: A11250; PMID:9832770; PMID:9665876
 A:Accession: G11299
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-682 <COU>
 A:Cross-references: GB:AE001239; GB:AE000520; NID:93322937; PIDN:AA65621.1; PID:9332294
 A:Experimental source: strain Nichols
 C:Genetics:
 A:Gene: TP0648

C:Superfamily: hypothetical protein TP0648
 Query Match 9.5%; Score 99; DB 1; Length 682;
 Best Local Similarity 25.2%; Pred. No. 1.7;
 Matches 41; Conservative 29; Mismatches 59; Indels 34; Gaps 6;

Qy 41 AISLDMGKLDQAKQOLDALSDROFAPAYRTLAKVQASEDATHQAKRFE 95
 Db 120 AFSLVGQGEELDQASLFEILARYPNVDARFGLAEIVSGRLSS-----ARLYQ 171
 Qy 96 KAIELNPKMOSYMDYGFYLVQMGDSLGLIYDQKSPRAIGYGRVALENMAYIYHQY 155
 Db 172 AALEROAENRKAALLSLALISYAGHYPRALTYVE---RALQYHGDNAQV-----HFF 220
 Qy 156 EAAKSPKDDYNNAKSALERALLSGTQHDRIKKSVDK---LLS 195
 Db 221 AAYATTLTAHYEDERLYERAL-----HIKSAVPRRALLS 256

RESULT 34
 F87263
 TPR domain protein [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: F87263
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Neilson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, U.J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; PMID:21173698; PMID:11259647
 A:Accession: F87263
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-747 <STO>
 A:Cross-references: GB:AE005673; NID:913421226; PIDN:AAK22106.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0119

Query Match 9.5%; Score 99; DB 2; Length 747;
 Best Local Similarity 20.8%; Pred. No. 1.9;
 Matches 37; Conservative 33; Mismatches 70; Indels 38; Gaps 4;

Qy 30 NPOLAQRTOIAISLDMGKLDQAKQOLDALSDROFAPAYRTLAKVQASE 82
 Db 187 NPQAEFTKNAIILLKRAEBAACRSALALAPDPQVHLITGNALIDLRAAEAE 246
 Qy 83 D-----ATHQKAORLEFKAIELNPKMOSYMDYGFYLVQ 118
 Db 247 SFRQAIRLKPDYSEAHGNCALKLSGRITAEFCFRRAIQDANPADQAHNNLGDVFDL 306
 Qy 119 GDLGSLIYDQKSPRAIGYGRVALENMAYIYHQYEAKSPK---KDDYNNAKSA 172
 Db 307 GRFADAEAFY---RAAIGLKEPYLEAHNSNLLCLINYEFTSPDYTLAEAKQYGVASA 361

RESULT 35
 H69978
 conserved hypothetical protein yrbB - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jan-2000
 C:Accession: H69978
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beter
 C.; Bron, S.; Brunielle, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
 A.; Lathrop, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
 A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror

THIS PAGE BLANK (USPTO)

Qy 61LSADRFPAAYRFLAKYQASSEDATQHTKQRFEEKALIEINPDMQSYMDYGFYLVQMGD 12

Db 61 LGHDKNYIVLHSAHLAHYHQOQSGIEN---AFREYELAVNLNHQGDYHNNFGFPLCSQK 117

Qy 121 LSGALIFPDKPSRAIGREGRVALENMAYIY 152

Db 118 FEAOQOQFELALNPSNYHQADTFENIVLCAY 149

RESULT 2

OGT1_HUMAN

ID OGT1_HUMAN STANDARD; PRT; 1036 AA.

AC 015294; 096CC1;

DT 30-MAY-2000 (Rel. 39, Created)

DT 15-JUN-2002 (Rel. 41, last sequence update)

DT 15-JUN-2002 (Rel. 41, last annotation update)

DE UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110

DE kDa subunit (EC 2.4.4.1.-) (O-GlcNAc transferase p110 subunit).

GN OGT.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 2), AND SEQUENCE OF 217-226 AND 945-961.

RC TISSUE=Liver;

RX MEDLINE=97238670; PubMed=9083068;

RA Lubas W.A., Frank D.W., Krause M., Hanover J.A.;

RT "O-linked GlcNAc transferase is a conserved nucleocytoplasmic protein

RT containing tetratricopeptide repeats.";

RL J. Biol. Chem. 272:9316-9324(1997).

[2]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RX MEDLINE=21635556; PubMed=11773972;

RA Nolte D., Muller U.;

RT "Human O-GlcNAc transferase (OGT): genomic structure, analysis of

RT splice variants, fine mapping in Xq13.1.";

RL Mamm. Genome 13:62-64(2002).

[3]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RN TISSUE=Colon;

RC Strusberg R.;

RA Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.

CC -1- FUNCTION: ADDITION OF NUCLEOTIDE-ACTIVATED SUGARS DIRECTLY ONTO

CC THE POLYPEPTIDE THROUGH O-GLYCOSIDIC LINKAGE WITH THE HYDROXYL OF

CC SERINE OR THREONINE.

CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + peptide = UDP +

CC N-acetyl-beta-D-glucosaminyl-peptide.

CC -1- PATHWAY: glycosylation.

CC -1- SUBUNIT: HETEROOLIGOMER OF TWO 11 kDa AND ONE 70 kDa SUBUNITS. IT

CC IS NOT KNOWN IF THE 70 kDa SUBUNIT IS ENCODED BY A SEPARATE GENE

CC OR IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN

CC ALTERNATIVE INITIATION OF THE 110 kDa SUBUNIT (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC (POSSIBLE).

CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are

CC produced by alternative splicing.

CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS AND TO A LESSER

CC EXTENT IN SKELETAL MUSCLE, HEART, BRAIN AND PLACENTA. PRESENT IN

CC TRACE AMOUNTS IN LUNG AND LIVER.

CC -1- SIMILARITY: CONTAINS 12 TPR REPEATS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U77413; AAB61466.1; -

DR EMBL: AJ315767; CAC86127.1; -

DR EMBL: AJ315767; CAC86129.1; -

DR EMBL: BC014434; AAH14434.1; -

DR Genew; HGNC:8127; OGT.
DR MIM: 300255; -.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 8.
DR SMART; SM00028; TPR; 8.
DR KX transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat;
KW Alternative splicing.
FT REPEAT 79 112 TPR 1.
FT REPEAT 113 146 TPR 2.
FT REPEAT 147 180 TPR 3.
FT REPEAT 181 214 TPR 4.
FT REPEAT 215 248 TPR 5.
FT REPEAT 249 282 TPR 6.
FT REPEAT 317 350 TPR 7.
FT REPEAT 351 384 TPR 8.
FT REPEAT 385 418 TPR 9.
FT REPEAT 419 452 TPR 10.
FT REPEAT 453 463 TPR 11.
FT REPEAT 477 493 TPR 12 (INCOMPLETE).
FT DOMAIN 1 166 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT VARSPLIC 1 166 MASSGVNVAADSTGLAEHLAREYQAGDEFAARHCMLWROE
FT PFTSPHLSLTTP (IN ISOFORM 2).
SQ SEQUENCE 1036 AA; 115705 MW; C3BD67340925A2C CRC64;
Query Match 11.3%; Score 117.5; DB 1; Length 1036;
Best Local Similarity 25.5%; Pred. No. 0.029;
Matches 48; Conservative 33; Mismatches 68; Indels 39; Gaps 77

QY 10 VMANGVLVASCSPTPEPKNNPQLAIRTQIAISLDMGKIDQAKQOOLDAALSADROPAP 69
Db 123 LVAAQDEMGAVQAVYVALQYNNPDLYCVSRSDIGNLTKALGRLEBAKACYLKALETQPNFVAV 182
QY 70 AYRLRAYVQASED--ATTQTKQRLPEKALIELPKDMQSYMYQGFYLVQMGDSLGLI 126
Db 183 AMSNLGCVFNQGGIMWLAIH----FEKAVTIDPNFLDAVINLGNVLKS-----ARI 230
QY 127 YFDKPSRAIGYEGRVAIE-----NMAYIYHQ-----YEAA--KSPTKDY 166
Db 231 F---DRAVAAYIALSLSPHVAHVHGNLACVYEEQGLIDLAIDTYRRAIELQHPFPAV 286
QY 167 MNKASALE 174
Db 287 CNLNAALK 294

RESULT 3
OGTL RAT STANDARD; PRT; 1036 AA.
ID _OGTL RAT STANDARD; PRT; 1036 AA.
AC P6558; RAT
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase 110
DE kda subunit (EC 2.4.1.-) (O-GlcNAc transferase p110 subunit).
GN OGT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA MEDLINE=97238869; PubMed=9083067;
RA Kreppel L.K., Blomberg M.A., Hart G.W.;
RT "dynamic glycosylation of nuclear and cytosolic proteins. Cloning and
RT characterization of a unique O-GlcNAc transferase with multiple
RT tetratricopeptide repeats".
RT J. Biol. Chem. 272:19308-19315(1997).
CC -!- FUNCTION: ADDITION OF NUCLEOTIDE-ACTIVATED SUGARS DIRECTLY ONTO

CC THE POLYPEPTIDE THROUGH O-GLYCOSIDIC LINKAGE WITH THE HYDROXYL OF
 CC SERINE OR THREONINE.
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + peptide = UDP +
 CC N-acetyl-beta-D-glucosaminyl-peptide.
 CC -1- ENZYME REGULATION: BY TYROSINE PHOSPHORYLATION AND O-GLCNAC
 CC MODIFICATIONS.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBUNIT: HETERODIMER OF TWO 110 kDa AND ONE 78 kDa SUBUNITS. IT
 CC IS NOT KNOWN IF THE 78 kDa SUBUNIT IS ENCODED BY A SEPARATE GENE
 CC OR IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN
 CC ALTERNATIVE INITIATION OF THE 110 kDa SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOSOLIC (POSSIBLE).
 CC -1- TISSUE SPECIFICITY: APPEARS TO BE PRESENT IN ALL TISSUES EXAMINED
 CC EXCEPT KIDNEY.
 CC -1- PTM: AUTOMODIFIED BY O-GLYCOSYLATION WITH O-GLCNAC.
 CC -1- SIMILARITY: CONTAINS 12 TPR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, U76557; AAC53121.1; -
 CC InterPro: IPR001440; TPR.
 CC Pfam: PF00515; TPR; 12.
 CC SMART, SM00028; TPR; 11.
 CC DR Transferrase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat;
 CC KW Phosphorylation; Glycoprotein.
 CC FT REPEAT 79 112 TPR 1.
 CC FT REPEAT 113 146 TPR 2.
 CC FT REPEAT 147 180 TPR 3.
 CC FT REPEAT 181 214 TPR 4.
 CC FT REPEAT 215 248 TPR 5.
 CC FT REPEAT 249 282 TPR 6.
 CC FT REPEAT 283 316 TPR 7.
 CC FT REPEAT 317 350 TPR 8.
 CC FT REPEAT 351 384 TPR 9.
 CC FT REPEAT 385 418 TPR 10.
 CC FT REPEAT 419 452 TPR 11.
 CC FT REPEAT 453 483 TPR 12 (INCOMPLETE).
 CC FT DOMAIN 478 493 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC FT MOD RES 979 979 PHOSPHORYLATION (PROBABLE).
 CC SQ SEQUENCE 1036 AA; 115605 MW; 3F057CABDD0198D6 CRC64;
 CC -----
 CC Query Match 11.3%; Score 117.5; DB 1; Length 1036;
 CC Best Local Similarity 25.5%; Pred. No. 0.029;
 CC Matches 48; Conservative 33; Mismatches 68; Indels 39; Gaps 7;
 CC -----
 CC Oy 10 VMAMGLVLSACOSTPIPKNNPOLAQTOTAIISLIDNGKLDQAKQDADALSADROPAP 69
 CC Db 123 LVAAGDMGAAVQAVSALQVNDLYCVASDIGNLKLKALREBAKCYLKIETOPNFAY 182
 CC Oy 70 AVRTIAKYQASD---ATHQTKAQRLPEKAIELPKMOSYMDVGFYLVQMGDLSGALI 126
 CC Db 183 AMSNLGCTFPAKOGELWLAIH-----PEKAVTLDPNLDVAINIGNVLR-----ARI 230
 CC Oy 127 YFDRPSRAIGYGRVAIE-----NMAYIYHQ-----YEAA---KSPTKDY 166
 CC Db 231 F---DRAVAAYLRALSLSPNHAUVHGLACVYEGQIDLAIDYRRAIELQHPFPAY 286
 CC Oy 167 NNAKSALE 174
 CC Db 287 CNLANAK 294
 CC -----
 CC RESULT 4
 CC CC27_YEAST STANDARD; PRT; 758 AA.
 CC ID_CC27_YEAST P38042;
 CC DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cell division control protein 27.
 GN CDC27 OR SNB1 OR YBH084C OR YBL0718.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=96076635; PubMed=7502586;
 RA Obermaler B., Gassenhuber J., Piravandi E., Downey H.;
 RT "Sequence analysis of a 78.6 kb segment of the left end of
 RT Saccharomyces cerevisiae chromosome II.";
 RL Yeast 11:1103-1112(1995).
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A., AND MUTAGENESIS.
 RX MEDLINE=92306611; PubMed=1819514;
 RA Sikorski R.S., Michaud W.A., Woollon J.C., Boguski M.S., Connelly C.,
 RA Hieter P.A.;
 RT "TPR proteins as essential components of the yeast cell cycle.";
 RL Cold Spring Harb. Symp. Quant. Biol. 56:663-673(1991).
 RN [3]
 RP SUBUNITS.
 RX MEDLINE=95009933; PubMed=7925276;
 RA Lamb J.R., Michaud W.A., Sikorski R.S., Hieter P.A.;
 RT "Cdc16p, Cdc23p and Cdc27p form a complex essential for mitosis.";
 RL EMBO J. 13:4321-4328(1994).
 CC -1- FUNCTION: EXECUTES ESSENTIAL MITOTIC FUNCTIONS NEAR THE
 CC METAPHASE/ANAPHASE TRANSITION.
 CC -1- SUBUNIT: CDC16, CDC23 AND CDC27 FORMS A MACROMOLECULAR COMPLEX.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE CDC27/NUC2 FAMILY.
 CC -1- SIMILARITY: CONTAINS 8 TPR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X79489; CA556022.1; ALT_INIT.
 CC EMBL, Z35845; CA84905.1; -
 CC DR PIR; S45825;
 CC DR SGD; S0000180; CDC27.
 CC DR InterPro: IPR001440; TPR.
 CC DR Pfam: PF00515; TPR; 8.
 CC DR SMART, SM00028; TPR; 5.
 CC KW Cell division; Cell cycle; Mitosis; Repeat; TPR repeat;
 CC KW Nuclear protein.
 CC FT REPEAT 154 187 TPR 1.
 CC FT REPEAT 187 219 TPR 2.
 CC FT REPEAT 219 252 TPR 3.
 CC FT REPEAT 252 285 TPR 4.
 CC FT REPEAT 285 318 TPR 5.
 CC FT REPEAT 318 351 TPR 6.
 CC FT REPEAT 351 384 TPR 7.
 CC FT REPEAT 384 417 TPR 8.
 CC FT DOMAIN 418 433 ASN-RICH.
 CC FT MUTAGEN G-D: IN TEMPERATURE SENSITIVE MUTANT.
 CC SQ SEQUENCE 758 AA; 85436 MW; 8612E4504337A02 CRC64;
 CC -----
 CC Query Match 10.8%; Score 113; DB 1; Length 758;
 CC Best Local Similarity 23.8%; Pred. No. 0.048;
 CC Matches 46; Conservative 43; Mismatches 66; Indels 38; Gaps 9;
 CC -----
 CC Oy 38 TQIAISLD-----MGKLDQAKQDADALS-----DROFAPAYRTIAKYQ 79
 CC Db 526 SNLNLGMDTPNKPRTWCCTGNNLSLQKHDAIKAEKATQDLPNFAYAY-TLOGHEH 584

```

QY 80 ASEDATHOTKAOPLFEKATLEPNKDMQSYMDYGYLVOMGDLGALLIYFDKPSRAIGYEG 139
DB 585 SSNDSDSAKT--CYRKALACDPQHNAYAGLTGSAMKGGQYEALLEYEK-ARSLNPNV 641
QY 140 RVV-----AIENMAY-----IYHOYEAKSPKDDYNNAKSALBERALISGTHDEIKS 189
DB 642 VYLICCGGSLLEKLGKRYELACHLOPTS---SLSKYMGGLVLSMRYNVALQT 698
QY 190 YDKLSDYKLLSD 202
DB 699 FELV---KLIVPD 708

RESULT 5
Y798.METUA STANDARD; PRT; 334 AA.
ID Y798.METUA
AC 058208;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0798.
GN MJ0798.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_Taxid=2190;
[1]
SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uettermann T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts J.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RT Science 273:1058-1073 (1996).
RL -1- SIMILARITY: CONTAINS 7 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U67524; AAB98793.1; -.
CC DR TIGR; M0798; -.
CC DR InterPro; IPR001440; TPR.
CC DR Pfam; PF00515; TPR; 6.
CC DR SMART; SM00028; TPR; 5.
CC KW Hypothetical protein; Repeat; TPR repeat; Complete proteome.
FT REPEAT 102
FT REPEAT 135
FT REPEAT 137
FT REPEAT 168
FT REPEAT 169
FT REPEAT 202
FT REPEAT 204
FT REPEAT 235
FT REPEAT 236
FT REPEAT 269
FT REPEAT 273
FT REPEAT 306
FT REPEAT 308
FT REPEAT 333
SQ SEQUENCE 334 AA; 39695 MW; 1FC759BFI5E64970 CRC64;

Query Match
Best Local Similarity 9.7%; Score 101.5; DB 1; Length 334;
Matches 40; Conservative 29; Mismatches 68; Indels 31; Gaps 5;
QY 43 SLIDMGKLDQAKQDLALSDRQFAPAYRTLAKYQASDATHOTKAOPLFEKATLEPN 102
DB 144 ALLKLYKDLAIKFEKASEKDRNNYKALFGKSYILMSDMKNSIK--YEKYLELNP 200

```

```

QY 103 KDMQSYMDYGYLVOMGDLGALLIYFDKPSRAIGYEGVAIE-----NMATYYHQ 154
DB 201 NDVEA-LEY-----LGEIYYEEDCEKAINVFKALIKDPDIDLILKVAFTYK 249
QY 155 YEAKSPKDDYNNAKSALBERALISGTHDEIKSYDKLSDYKLLSD 202
DB 250 KK-----YKHALKTFEYALKLNPNVFELEOYESMGRIYYLGE 288

RESULT 6
MYSD CAEEL STANDARD; PRT; 1938 AA.
ID MYSD CAEEL
AC P02567; Q19674;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain D (MHC D).
GN MYO-1 OR R06C7.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=89178677; PubMed=6576820;
RA Dobb N.J., Maruyama I.N., Krause M., Karn J.;
RT "Sequence analysis of the complete Caenorhabditis elegans myosin
RT heavy chain gene family."
RT J. Mol. Biol. 205:603-613 (1989).
[2]
SEQUENCE OF 34-1795 FROM N.A.
RX MEDLINE=83273600; PubMed=6576334;
RA Karn J., Brenner S., Barnett L.;
RT "Protein structural domains in the Caenorhabditis elegans unc-54
RT myosin heavy chain gene are not separated by introns."
RT Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257 (1983).
[3]
SEQUENCE OF 115-365 AND 1492-1763 FROM N.A.
RX MEDLINE=85201409; PubMed=388374;
RA Karn J., Dobb N.J., Miller D.M.;
RT "Cloning nematode myosin genes."
RT Cell Muscle Motil. 6:185-237 (1985).
[4]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Gardner A., McMurtry A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils;
CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
CC C.ELEGANS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL, X08065; CAA30854.1; -
 CC or send an email to license@isb-sib.ch.
 DR EMBL, M37232; AAA28119.1; -
 DR EMBL, M37234; AAA28120.1; -
 DR EMBL, 271266; CAA95848.1; -
 DR EMBL, 271261; CAA95848.1; JOINED.
 DR EMBL, 271261; CAA95806.1; -
 DR EMBL, 271266; CAA95806.1; JOINED.
 DR PIR, S02772; MKKML.
 DR HSBP, P08799; LMND.
 DR WormPep; R06C7.10; CE06253.
 DR InterPro; IPR004009; Myosin.N.
 DR InterPro; IPR002928; Myosin.N.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SMO0242; MYSC; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding; ATP-binding; Methylation; Alkylation; Multigene family.
 KW DOMAIN 1 845 MYOSIN HEAD-LIKE.
 FT DOMAIN 846 1938 RODLIKE TAIL (S2 AND LMM DOMAINS).
 FT DOMAIN 846 1170 ALPHA-HELICAL TAILPIECE (SHORT S2).
 FT DOMAIN 1171 1938 LIGHT MEROMYOSIN (LMM).
 FT DOMAIN 846 1938 COILED COIL (POTENTIAL).
 FT NP_BIND 177 184 ATP (BY SIMILARITY).
 FT DOMAIN 660 682 ACTIN-BINDING.
 FT DOMAIN 764 778 ACTIN-BINDING.
 FT MOD_RES 128 128 METHYLATION (SH-1).
 FT MOD_RES 700 700 ALKYLATION (SH-2).
 FT MOD_RES 710 710 ALKYLATION (SH-2).
 FT CONFLICT 94 94 F -> E (IN REF. 2).
 FT CONFLICT 98 98 A -> R (IN REF. 2).
 FT CONFLICT 377 377 V -> D (IN REF. 4).
 FT CONFLICT 389 390 DV -> GD (IN REF. 2).
 FT CONFLICT 391 391 V -> D (IN REF. 4).
 FT CONFLICT 408 408 W -> N (IN REF. 2).
 FT CONFLICT 474 474 Q -> G (IN REF. 2).
 FT CONFLICT 577 577 L -> F (IN REF. 4).
 FT CONFLICT 681 681 I -> N (IN REF. 4).
 FT CONFLICT 1373 1373 S -> D (IN REF. 2).
 FT CONFLICT 1659 1659 E -> Q (IN REF. 3).
 SO SEQUENCE 1938 AA; 223255 MW; 387399C8F63A4CF4 CRC64;

Query Match 9.5%; Score 99.5; DB 1; Length 1938;
 Best Local Similarity 27.7%; Pred. No. 2.1;
 Matches 49; Conservative 34; Mismatches 65; Indels 29; Gaps 9;

QY 29 NNPOLAQRTOIAISL-LDMGKLDQAQOOLDALSL-ADROPAPAYRTAKYVQASEDAATH 86
 DB 1726 NNAALISAMKRVKENEVOIARBELDEYNELKASERAKAADAADRLAEVYQOEGBAHVH 1785
 QY 87 QTKAQRLEKALINPKMOSYMDGYFLVQMGDLSGALTYFDKKSRAIGEGRAVALEN 146
 DB 1786 VDR-OR---KSELELAKELQAKID-----DERRMIDPGAKVALA-KVEDRRSLR- 1830
 QY 147 MAYIYVHO---YEAAKSPTKDDYNNAKSALERALISGTQHDIEIKKSYDKLSDYQL 200
 DB 1831 -AELHSEGRRHQESIKGYTKOE-----RRARLQGVVEDKKAFLRLQENVEKL 1878

RESULT 7
 OM70_NEUCR STANDARD; PRT; 619 AA.
 AC P23231;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mitochondrial precursor proteins import receptor (72 kDa mitochondrial

outer membrane protein) (Mitochondrial import receptor for the ADP/ATP carrier) (Translocase of outer membrane TOM70).
 GN TOM70 OR MOM72.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A;
 RX MEDLINE=91115930; PubMed=2177474;
 RA Steiger H.F., Soellner T., Kiebler M., Dietmeier K.A., Pfaller R.,
 RA Trueilsch K.S., Tropschug M., Neupert W., Pfanner N.;
 RT "Import of ADP/ATP carrier into mitochondria: two receptors act in parallel";
 RL J. Cell Biol. 111:2353-2363 (1990).
 CC -1- FUNCTION: RECEPTOR THAT ACCELERATES THE IMPORT OF ALL
 CC MITOCHONDRIAL PRECURSOR PROTEINS. SEEMS TO ACT IN CONJUNCTION WITH
 CC MOMO19.
 CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
 CC LEAST SEVEN DIFFERENT PROTEINS (MOM1, MOM6, MOM19, MOM22, MOM30,
 CC MOM38 AND MOM72).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC outer membrane.
 CC -1- SIMILARITY: CONTAINS 11 TPR REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, X53735; CAA37767.1; -
 DR PIR; A36682; A36682.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 9.
 DR SMART; SMO0028; TPR; 9.
 DR TIGRFAMs; TIGR00990; 360801809; 1.
 KW Mitochondrion; Outer membrane; Transmembrane; TPR repeat; Receptor;
 KW Repeat.
 FT TRANSMEM 39 60 POTENTIAL.
 FT DOMAIN 61 619 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 131 164 TPR 1.
 FT REPEAT 166 197 TPR 2.
 FT REPEAT 199 231 TPR 3.
 FT REPEAT 297 331 TPR 4.
 FT REPEAT 335 368 TPR 5.
 FT REPEAT 370 402 TPR 6.
 FT REPEAT 403 436 TPR 7.
 FT REPEAT 438 470 TPR 8.
 FT REPEAT 471 504 TPR 9.
 FT REPEAT 512 545 TPR 10.
 FT REPEAT 546 579 TPR 11.
 SO SEQUENCE 619 AA; 68917 MW; 41FEB932B8F8E1F0 CRC64;

Query Match 9.3%; Score 97.5; DB 1; Length 619;
 Best Local Similarity 21.8%; Pred. No. 0.78;
 Matches 37; Conservative 36; Mismatches 78; Indels 17; Gaps 4;

QY 13 MGVLISACQSTPIPPKNNPOLAQRTOIAISLDMGKLDQAQOOLDALSLADROPAP- 69
 DB 450 MGSIASSMAITFRCKMFPDQTPDYNYNGELLIDQNFQEAIEKFDTRIALEKSTKPMCM 509
 QY 70 -----AVRTLAKYVQASEDAATHQTKAQRLEKALINPKMOSYMDGYFLVQMGDLSGA 124
 DB 510 NVLPLINKALA-LFQWQD---YAEAQOLEKRLIIPBEDINAVTMAQILLQGGKYVEA 565
 QY 125 LIYDKFSRAIGEGRAVALENNAVITYYHQEYEAASPTKDDYNNAKSALE 174
 DB 566 LKFEPAEAELARTEGELVNA-----LSYAEATRTQIOVQENYBELASKIQ 610

```

RESULT 8
YD45 META ID YD45 META STANDARD; PRT; 314 AA.
AC 058741;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein MJ1345.
GN MJ1345.
OS Methanococcus jannaschii;
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxId=2190;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,
RA Uhlirback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurest M.A., Kaine B.P., Borodovsky M.,
RA Kleink H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
RN 2;
RP DISCUSSION ON SEQUENCE.
RX MEDLINE=98362752; PubMed=9697413;
RA Kyriides N.C., Woese C.R.;
RT "Tetratric-peptide-repeat proteins in the archaeon Methanococcus
RL jannaschii.";
RN Trends Biochem. Sci. 23:245-247(1998).
CC -I- SIMILARITY: CONTAINS 9 TPR REPEATS.
-----
CC This SMSS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; U67574; AAB9354.1; -.
DR TIGR; MJ1345; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 8.
DR SMART; SMO0028; TPR; 7.
KW Hypothetical protein; Repeat; TPR repeat; Complete proteome.
FT REPEAT 12 45 TPR 1.
FT REPEAT 46 78 TPR 2.
FT REPEAT 80 112 TPR 3.
FT REPEAT 114 146 TPR 4.
FT REPEAT 147 180 TPR 5.
FT REPEAT 182 214 TPR 6.
FT REPEAT 215 248 TPR 7.
FT REPEAT 249 282 TPR 8.
FT REPEAT 284 313 TPR 9.
SQ SEQUENCE 314 AA; 36518 MW; F58D35315E854E12 CRC64;

```

```

Query Match      9.2%; Score 96; DB 1; Length 314;
Best Local Similarity 21.7%; Pred.No. 0.46;
Matches 36; Conservative 25; Mismatches 53; Indels 52; Gaps 4

QY      48 GKLDPAKQQQLDAALSAD-----RQFAPAVRYTLAKVYAQSADATHQ--- 87
          |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       161 GKLRKSLCEFDNALINIPKDCOSLLYKGTEIFLGRGYSBALCKLKCFERNNNKDRIALMY 220
          ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      88 -----TKAQRLEPKATLELNPKMQSGTMDGGFYLVQMGDSGLAIYEDK-----P 131
          :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Db 221 IIQILLYFSLNQALEYETKKAFLNPDDPLLYLVYKGIILNTLGRYNEAIKIFDVLVEINP 280
 QY 132 SRALGEGRGVALENNAYLYYHQYEAASPTKDDYNNAKSLBEALL 177
 Db 281 NIPDANNKGAIALEKTKGI-----NEAIECYNAL 310

	RESULT 9		
ID	SOLR_CLOAB	STANDARD:	PRT; 318 AA.
AC	P33746;		
DT	01-FEB-1994 (rel. 28, Created)		
DT	01-FEB-1994 (rel. 28, Last sequence update)		
DT	15-JUN-2002 (rel. 41, Last annotation update)		
DE	Sol locus transcriptional repressor.		
GN	SOLR OR CAP0161.		
OS	Clostridium acetobutylicum.		
OC	Plasmid pSOL1.		
CC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
CC	Clostridium.		
OX	NCBI_TaxID=1488;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 824 / DSM 792 / VKM B-1787;		
RX	MEDLINE=94042861; PubMed=8226639;		
RA	Fischer R.J., Helms J., Duarte P.;		
RT	Cloning, sequencing, and molecular analysis of the sol operon of		
RT	Clostridium acetobutylicum, a chromosomal locus involved in		
RL	solventogenesis.";		
RT	J. Bacteriol. 175:6959--6969(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 824 / DSM 792 / VKM B-1787;		
RX	MEDLINE=99084972; PubMed=986345;		
RA	Nair R.V., Green E.M., Watson D.E., Bennett G.N., Papoutsakis E.T.;		
RT	"Regulation of the sol locus genes for butanol and acetone formation		
RT	in Clostridium acetobutylicum ATCC 824 by a putative transcriptional		
RT	repressor.";		
RL	J. Bacteriol. 181:319-330(1999).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 824 / DSM 792 / VKM B-1787;		
RX	MEDLINE=21359325; PubMed=11466286;		
RA	Noelling J., Brexon G., Omelchenko M.V., Makarova K.S., Zeng Q.,		
RA	Gibson R., Lee H.M., Dubois J., Qiu D., Hitt J., Wolf Y.I.,		
RA	Tatusov R.V., Sabatne F., Doucette-Stamm L., Soucaille P., Daly M.J.,		
RA	Bennett G.N., Koonin E.V., Smith D.R.;		
RT	"Genome sequence and comparative analysis of the solvent-producing		
RT	bacterium Clostridium acetobutylicum".		
RT	J. Bacteriol. 183:4823-4838(2001).		
CC	-I- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF THE SOL LOCUS (ADHE/AAD,		
CC	CTPA, CTPB AND ADC) GENES FOR BUTANOL AND ACETONE FORMATION.		
CC	-I- SIMILARITY: CONTAINS 4 TPR REPEATS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to licence@isb-sib.ch).		
CC	-----		
CC	EMBL; X72831; CAAS1342.1; -		
CC	EMBL; L14817; AADO4637.1; -		
CC	EMBL; AE001438; AAK76906.1; -		
CC	PIR; S33433; S33433;		
CC	InterPro; IPR001440; TPR.		
CC	pfam; PF00515; TPR; 5.		
CC	SMART; SMOO028; TPR; 4.		
KW	Transcription regulation; Repressor; Repeat; TPR repeat; Plasmid;		
KW	Complete proteome.		
FT	REPEAT 65 98 TPR 1.		
FT	REPEAT 99 132 TPR 2.		

FT REPEAT 133 166 TPR 3.
 FT REPEAT 167 199 TPR 4.
 SQ SEQUENCE 318 AA; 36939 MW; 49CED287FE4EF8 CRC64;

Query Match 9.2%; Score 96; DB 1; Length 318;
 Best Local Similarity 25.9%; Pred. No. 0.47;

Matches 38; Conservative 23; Mismatches 52; Indels 34; Gaps 7;

QY 68 APAYRTLLKAVYQASDATHQTVARQLFE-----KATLELNPKDMQSYMDYGFYLVOMGD 120
 DB 65 ANAYITRIRIYFYSND-----FELSLRDLQAIKLRPKTINDYSPALSTHILGE 114
 QY 121 LSGALIVYDKPSRAIGYGRV-VALENNAYIYHQYEAKSPTKDYNNAKALERALIS 179
 DB 115 PERALKYF--LRAYELQPNVIGISYENLAMPFY-----LTGKYDKAIENFEKALISM 162
 QY 180 GTQHDEIKK---SYDKLSDYKLSDY 203
 DB 163 GSTNSVYSRLGITVAK-IGDYKKSSEY 188

RESULT 10

KEAP_HUMAN STANDARD; PRT; 624 AA.

AC Q14145;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Kelch-like ECH-associated protein 1 (Cytosolic inhibitor of Nrf2).
 GN KEAP1 OR KIA0132.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP TISSUE=Bone marrow;
 RC MEDLINE=96127530; PubMed=8559280;
 RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;

RT "Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (K1A0121-K1A0160) deduced by analysis of cDNA clones from human cell line KG-1."
 RL DNA Rep. 2:167-174(1995).
 CC -1- FUNCTION: RETAINS NFE2L2/NRF2 IN THE CYTOSOL THUS RESULTING IN THE SUPPRESSION OF ITS TRANSCRIPTIONAL ACTIVITY AND THE REPRESSION OF ANTIOXIDANT RESPONSE ELEMENT-MEDIATED DETOXIFYING ENZYME GENE EXPRESSION (BY SIMILARITY).
 CC -1- SUBUNIT: INTERACTS WITH THE N-TERMINAL REGULATORY DOMAIN OF NFE2L2/NRF2 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, D50922; BAA09481.1; -.
 DR MIM; 606016; -.
 DR InterPro; IPR000310; BTB_POZ.
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF01344; Kelch; 6.
 DR PRINTS; PR00501; KELCHREPEAT.
 DR SMART; SM00225; BTB; 1.
 DR PROSITE; PS50097; BTB; 1.
 KM Transcription regulation; Repeat.
 FT DOMAIN 77 149 BTB.
 FT REPEAT 327 372 KELCH 1.

FT REPEAT 373 423 KELCH 2.
 FT REPEAT 424 470 KELCH 3.
 FT REPEAT 471 517 KELCH 4.
 FT REPEAT 518 564 KELCH 5.
 FT REPEAT 565 611 KELCH 6.
 SQ SEQUENCE 624 AA; 69665 MW; 7052EF3BF436C90 CRC64;

Query Match 9.1%; Score 94.5; DB 1; Length 624;
 Best Local Similarity 25.0%; Pred. No. 1.4;
 Matches 52; Conservative 32; Mismatches 57; Indels 67; Gaps 12;

QY 47 MGKLDQAKQQLDAALSDRQFAPAYRTLA-KVYQASDA-----THQTKAQLRFEKAIE 99
 DB 67 MNEIRLSQLQCDVTLQYKQDAPAPAFMAHKVVLASSSPVFKAMFTNGLRQGMWVSI 126
 QY 100 -LNFKDMQSYMDYGFYLVOMGD-----LSGALIV-FDK-----PSRAIG 136
 DB 127 GHPKRMVERLIEFAYTASISGKCVLHVMGAVMYOIDSVPKACDFLYVOQLDPSNAIG 186
 QY 137 YEGRVVAIENNA-----YIYVHQYEAASP-----TKDDYN- 167
 DB 187 -----IANPAEIGICVELHQRAREYIYMHFGVAKQBERFNLSHCOLVTLISRDLLN 239
 QY 168 NAKSALERALISGTQHD-EIKKSYDKLL 194
 DB 240 RCESEVFHACINWVKYDCEGRFRFYQAL 267

RESULT 11

YA88_AQUAE STANDARD; PRT; 761 AA.

AC O67178;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AQ_1088.
 GN AQ_1088.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;

CC Aquifex.
 OX NCBI_TaxId=63363;
 RN [1]
 RP STRAIN=VF5;
 RC MEDLINE=98196666; PubMed=9537320;
 RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham R.A., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus."
 RL Nature 392:353-358(1998).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO THE PROTEIN SULFOFRANSFERASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 6 TPR REPEATS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AE000722; AAC07141.1; -.
 DR InterPro; IPR000863; Sulfofransferase.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 4.
 DR Pfam; PF00685; Sulfofransfer; 1.
 DR SMART; SM00028; TPR; 4.
 KM Hypothetical protein; Repeat; TPR repeat; Complete proteome.
 FT REPEAT 35 68 TPR 1.
 FT REPEAT 69 102 TPR 2.
 FT REPEAT 103 136 TPR 3.


```

FT REPEAT 467 501 TPR 7.
FT REPEAT 502 535 TPR 8.
FT REPEAT 536 569 TPR 9.
FT REPEAT 569 598 TPR 10.
FT REPEAT 598 627 TPR 11.
FT REPEAT 627 656 TPR 12.
SQ SEQUENCE 1025 AA; 120228 MW; 12DE18304AAAF4662 CRC64;

Query Match
Best Local Similarity 21.5%; Pred. No. 4.6;
Matches 41; Conservative 29; Mismatches 82; Indels 39; Gaps 4;

QY 30 NPQLAQIRTOIAISLDMGKLDQAKQDALSADROFAPARYTLAKYQSEDATHQTK 79
DB 125 DPEVAQLLSQ-ANBAFVRNDLCVAAERLFNEVYIKDARNFPAAYETLGDYIQLGRINDCN 183
QY 80 -----ASEDATHTQTKAQLRLEKAIELNPKDMOSYMDYGFYLVQM 118
DB 184 SWFLAHLNASTDWEKWKVAILISADLDHVRQAIYCFSRVTSINPEWESIYRSMLYKKT 243
QY 119 GDLGALYFDKPSRAIGYGRVVAIENMAYIYHQYEAKSPTRDDYN-NAK-----S 171
DB 244 GQLARALDGFQRLVWYNPYDANILRELAIVLYDYDRIDISILYKMKVFANVERREALIA 303
QY 172 ALERLLISGTQ 182
DB 304 ALENALDSDE 314

RESULT 14
IFT1_HUMAN
ID IFT1_HUMAN STANDARD; PRT; 478 AA.
AC P09914; Q960M5;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Interferon-induced protein with tetratricopeptide repeats 1 (IFT1-1)
DE (Interferon-induced 56 kDa protein) (IFT1-56K).
GN IFT1 OR IFI56 OR GI0P1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RA [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86136112; PubMed=3753936;
RA Wathelet M., Mouschen S., Defilippi P., Cravador A., Collet M.,
RA Huez G., Content J.;
RT "Molecular cloning, full-length sequence and preliminary
RT characterization of a 56-kDa protein induced by human interferons.";
RL Eur. J. Biochem. 155:11-17(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 413-478 FROM N.A.
RX MEDLINE=83143342; PubMed=6186990;
RA Chebath J., Merlin G., Metz R., Benesh P., Revel M.;
RT "Interferon-induced 56,000 Mr protein and its mRNA in human cells:
RT molecular cloning and partial sequence of the cDNA.";
RL Nucleic Acids Res. 11:1213-1226(1983).
RN [4]
RP SEQUENCE OF 1-2 FROM N.A.
RX MEDLINE=88082760; PubMed=3121313;
RA Wathelet M.G., Clausen I.M., Nols C.B., Content J., Huez G.A.;
RT "New inducers revealed by the promoter sequence analysis of two
RT interferon-activated human genes.";
RL Eur. J. Biochem. 169:313-321(1987).
RN [5]
RP SIMILARITY TO IFT1-54K.
RX MEDLINE=88196376; PubMed=3350121;
RA Wathelet M.G., Clausen I.M., Content J., Huez G.A.;

```

```

RT "The IFT1-56K and IFT1-54K interferon-inducible human genes belong to
RT the same gene family.";
RL FEBS Lett. 231:164-171(1988).
CC -1- INDUCTION: BY INTERFERONS.
CC -1- SIMILARITY: BELONGS TO THE IFT1 FAMILY.
CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; X03557; CA27244.1; -.
DR EMBL; BC007091; AA07091.1; -.
DR EMBL; X06559; CA29802.1; -.
DR EMBL; M24594; AA59131.1; -.
DR PIR; A25407; A25407.
DR GenBank; HGNC:5407; IFT1.
DR MIM; 147690; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 6.
DR SMART; SM00028; TPR; 4.
DR Repeat; TPR repeat; Interferon induction.
FT REPEAT 52 85 TPR 1.
FT REPEAT 95 128 TPR 2.
FT REPEAT 139 174 TPR 3.
FT REPEAT 183 216 TPR 4.
FT REPEAT 218 249 TPR 5.
FT REPEAT 251 284 TPR 6.
FT REPEAT 305 339 TPR 7.
FT REPEAT 340 373 TPR 8.
FT REPEAT 378 412 TPR 9.
FT REPEAT 437 470 TPR 10.
FT CONFLICT 437 470 Y->H (IN REF. 2).
SQ SEQUENCE 478 AA; 55386 MW; EEDPCE308675731 CRC64;

Query Match
Best Local Similarity 22.3%; Pred. No. 2.5;
Matches 39; Conservative 28; Mismatches 78; Indels 30; Gaps 5;

QY 30 NPQLAQIRTOIAISLDMGKLDQAKQDALSADROFAPARYTLAKYQSEDATHQTK 89
DB 213 NPDNGYIKVLALQLQDGECAEGEKYIEELAMSSGTYYFRTAKYRKGSV---DK 269
QY 90 AQRLEKAIELNPKDMOSYMDYGF-YLVQMDLSGAL-----IYFDRPSRAIGYEG 139
DB 270 ALIELLKRALQETPFSVLLHQLGICVYKQMTQIKKATKGPRGQNRKLDGMRISA---- 325
QY 140 RVVAIENMAYIYHQYEAKSPTRDDYNNAKSALEKALISGTQDEIKKSYDKLL 194
DB 326 -----IFHESAVEKKPT--FEVALDLARMYIEAGNHRKAENFOKLL 367

RESULT 15
FRPA_NEIMC
ID FRPA_NEIMC STANDARD; PRT; 1115 AA.
AC P55126;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated protein ftrpa.
GN FRPA.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=135720;
RA [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FAM20 / Serogroup C;
RX MEDLINE=93139051; PubMed=8423153;
RA Thompson S.A., Wang L.L., West A., Sparling P.F.;

```

```

RT  "Neisseria meningitidis produces iron-regulated proteins related to
RT  the RTX family of exoproteins."
RT  J. Bacteriol. 175:811-818 (1993).
CC  -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
CC  DISEASE.
CC  -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED.
CC  -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC  CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC  ACTIVITY.
CC  -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, L06302; AAA25454.1; -.
DR  InterPro; IPR001343; HemIysn_Ca_bind.
DR  Pfam; PF00353; hemolysinCbind; 7.
DR  PRINTS; PR00313; CABNDNGRPT.
DR  PROSITE; PS00330; HEMOLYSIN CALCIUM; 5.
KW  Toxin; Calcium; Outer membrane; Repeat.
FT  DOMAIN 765 988 13 X REPEATS, GLY-RICH.
FT  REPEAT 774 779 2.
FT  REPEAT 783 788 3.
FT  REPEAT 902 907 4.
FT  REPEAT 911 916 5.
FT  REPEAT 920 925 6.
FT  REPEAT 929 934 7.
FT  REPEAT 938 943 8.
FT  REPEAT 947 952 9.
FT  REPEAT 956 961 10.
FT  REPEAT 965 970 11.
FT  REPEAT 974 979 12.
FT  REPEAT 983 988 13.
SQ  SEQUENCE 1115 AA; 122177 MW; 21841065BCACEA CRC64;

Query Match      8.6%; Score 90; DB 1; Length 1115;
Best Local Similarity 23.2%; Pred. No. 6.9;
Matches 58; Conservative 33; Mismatches 69; Indels 90; Gaps 15;

QY  1 MKIRVMPVMWANGVLVSACQSTPIPPKNNPQLAIRTOAISLDMK--LDQAKQ-- 56
DB  552 MRLSTWQTOTANEGIAL-----TP-----SQVAQLKKNALVSLDKAKAIDAARDRIA 600
QY  57 -LDAALASDRQFAPARYRLAKVQASEDA-----THQTKAQLFEKAI---ELNPK 103
DB  601 VLDAYTGQDS-----TL--YMSSEDAINIVKTNIDYDLAKNIYQNLFFQRLQP- 651
QY  104 DMQSYMDYGFYLVQMG-----DLGALIFYD-----KPSRAI----- 135
DB  652 -----YLNQISFKWENDFTLDPFSLVQAFNHVKEINQKAFVLDLAEMLAYGELR 701
QY  136 -GYEGVALENNAYVHYHOYEAQSPKDYDNNNAKSALEKALI---SGTGHDEIKS-- 189
DB  702 SWEGGRLLMAD-----YEEAKKAKGFEDYQKVLGOETVALLAKTSGTQADDIQNVG 754
QY  190 --YDKLLSDY 197
DB  755 FGHKNVSLY 764

RESULT 16
FRPA NEIMB STANDARD; PRT; 1302 AA.
AC  Q9K0K3;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Iron-regulated protein frpa.

```

```

GN  FRPA OR NMB0585.
OS  Neisseria meningitidis (serogroup B).
OC  Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX  NCBI_TaxID=491;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-MC58 / Serogroup B;
RX  MEDLINE=20175755; PubMed=10710307;
RA  Tetteh H., Saunders N.C.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA  Eissen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA  Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA  Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA  Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA  Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA  Gill J., Scariato V., Nasigant V., Pizsa M., Grandi G., Sun L.,
RA  Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venier J.C.;
RT  "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT  MC58."
RL  Science 287:1809-1815 (2000).
CC  -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
CC  DISEASE.
CC  -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED (BY
CC  SIMILARITY).
CC  -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC  CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC  ACTIVITY.
CC  -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, AB002414; AAF41013.1; -.
DR  TIGR; NMB0585; -.
DR  InterPro; IPR001343; HemIysn_Ca_bind.
DR  Pfam; PF00353; hemolysinCbind; 12.
DR  PRINTS; PR00313; CABNDNGRPT.
DR  PROSITE; PS00330; HEMOLYSIN CALCIUM; 9.
KW  Toxin; Calcium; Outer membrane; Repeat; Complete proteome.
FT  DOMAIN 752 975 13 X REPEATS, GLY-RICH.
FT  REPEAT 761 766 2.
FT  REPEAT 770 775 3.
FT  REPEAT 889 894 4.
FT  REPEAT 898 903 5.
FT  REPEAT 907 912 6.
FT  REPEAT 916 921 7.
FT  REPEAT 925 930 8.
FT  REPEAT 934 939 9.
FT  REPEAT 943 948 10.
FT  REPEAT 952 957 11.
FT  REPEAT 961 966 12.
FT  REPEAT 970 975 13.
SQ  SEQUENCE 1302 AA; 141397 MW; 21D058C56C98BDE8 CRC64;

Query Match      8.6%; Score 90; DB 1; Length 1302;
Best Local Similarity 23.2%; Pred. No. 8.3;
Matches 58; Conservative 33; Mismatches 69; Indels 90; Gaps 15;

QY  1 MKIRVMPVMWANGVLVSACQSTPIPPKNNPQLAIRTOAISLDMK--LDQAKQ-- 56
DB  539 MRLSTWQTOTANEGIAL-----TP-----SQVAQLKKNALVSLDKAKAIDAARDRIA 587
QY  57 -LDAALASDRQFAPARYRLAKVQASEDA-----THQTKAQLFEKAI---ELNPK 103
DB  588 VLDAYTGQDS-----NTL--YMSSEDAINIVKTNIDYDLAKNIYQNLFFQRLQP- 638
QY  104 DMQSYMDYGFYLVQMG-----DLGALIFYD-----KPSRAI----- 135
DB  639 -----YLNQISFKWENDFTLDPFSLVQAFNHVKEINQKAFVLDLAEMLAYGELR 688

```

Qy 136 -GYEGRVAIENMAYIYHOYEAKSPKDYNNAKSALERALI---SGTQHDEIKS-- 189
 Db 689 SWYEGRLMTD-----YVEAKKAGKGFEDYQKVLGOETVALLAKTSGTQADDLQNVG 741
 Qy 190 --YDKLSDY 197
 Db 742 FGHNKVSLY 751

RESULT 17
 FRPC_NEIMB STANDARD; PRT; 1829 AA.
 AC Q9JYV5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Iron-regulated protein frpc.
 GN FPRC OR NMB1415.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 Halt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Macon T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
 Ra Cotton M.D., Ueberback T.R., Khouli H., Qin H., Vamathevan J.,
 Gail J., Scaratti V., Maignan V., Pizzo M., Grandi G., Sun L.,
 Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venter J.C.,
 RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
 RT MCS8.";
 RL Science 287:1809-1815(2000).
 CC -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
 CC DISEASE.
 CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED (BY
 CC SIMILARITY).
 CC -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
 CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
 CC ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; AE002490; AAF1776.1; -
 DR TIGR; NMB1415; -
 DR InterPro; IPR001343; Hemlyan_Ca_bind.
 DR Pfam; PF00353; hemolysinCbind; 22.
 DR PRINTS; PR00313; CAENDNGRPT.
 DR PROSITE; PS00330; HEMOLYSIN_CALCITUM; 17.
 DR KX toxin; Outer membrane; Repeat; Complete proteome.
 FT DOMAIN 879 1702 43 X REPEATS, GLY-RICH.
 FT REPEAT 879 884 1.
 FT REPEAT 888 893 2.
 FT REPEAT 897 902 3.
 FT REPEAT 1016 1021 4.
 FT REPEAT 1025 1030 5.
 FT REPEAT 1034 1039 6.
 FT REPEAT 1043 1048 7.
 FT REPEAT 1052 1057 8.
 FT REPEAT 1061 1066 9.
 FT REPEAT 1070 1075 10.
 FT REPEAT 1079 1084 11.

FT REPEAT 1088 1093 12.
 FT REPEAT 1097 1102 13.
 FT REPEAT 1216 1221 14.
 FT REPEAT 1225 1230 15.
 FT REPEAT 1234 1239 16.
 FT REPEAT 1243 1248 17.
 FT REPEAT 1252 1257 18.
 FT REPEAT 1261 1266 19.
 FT REPEAT 1270 1275 20.
 FT REPEAT 1279 1284 21.
 FT REPEAT 1288 1293 22.
 FT REPEAT 1297 1302 23.
 FT REPEAT 1416 1421 24.
 FT REPEAT 1425 1430 25.
 FT REPEAT 1434 1439 26.
 FT REPEAT 1443 1448 27.
 FT REPEAT 1452 1457 28.
 FT REPEAT 1461 1466 29.
 FT REPEAT 1470 1475 30.
 FT REPEAT 1479 1484 31.
 FT REPEAT 1488 1493 32.
 FT REPEAT 1497 1502 33.
 FT REPEAT 1616 1621 34.
 FT REPEAT 1625 1630 35.
 FT REPEAT 1634 1639 36.
 FT REPEAT 1643 1648 37.
 FT REPEAT 1652 1657 38.
 FT REPEAT 1661 1666 39.
 FT REPEAT 1670 1675 40.
 FT REPEAT 1679 1684 41.
 FT REPEAT 1688 1693 42.
 FT REPEAT 1697 1702 43.
 SQ SEQUENCE 1829 AA, 197445 MW, 8F63506E1F6D9B40 CRC64;

Query Match 8.6%; Score 90; DB 1; Length 1829;
 Best Local Similarity 23.2%; Pred. No. 12;
 Matches 58; Conservative 33; Mismatches 69; Indels 90; Gaps 15;

Qy 1 MKIRKVMWVAMGLVLSACOSTPIPRKNNQALQIRQVLAISLDMCK--LDQAKQ-- 56
 Db 666 MRLSTDMQTANESIAL--TP--SQAQLKKNALVLSDKAKAIDAARDRIA 714
 Qy 57 -LDNALSDROFAPAYTLAKVQASEDA-----THQTKARLFKFAI---ELNPK 103
 Db 715 VLDAVYTGDS-----NTL--YVSEEDALNIYVNTDYLHAKNIYQNLFFQTRLP- 765
 Qy 104 DMQSYMDGYFLVQMG-----DISGALIVFD---KPSRAI----- 135
 Db 766 -----YLNQISFKMENDFTLDFSGLVQAFNHVKETNPQKAFVLAEMLAYGELR 815
 Qy 136 -GYEGRVAIENMAYIYHOYEAKSPKDYNNAKSALERALI---SGTQHDEIKS-- 189
 Db 816 SWYEGRLMTD-----YVEAKKAGKGFEDYQKVLGOETVALLAKTSGTQADDLQNVG 868
 Qy 190 --YDKLSDY 197
 Db 869 FGHNKVSLY 878

RESULT 18
 FRPC_NEIMC STANDARD; PRT; 1829 AA.
 AC P55127;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Iron-regulated protein frpc.
 GN FPRC.
 OS Neisseria meningitidis (serogroup C).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=135720;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=PAM20 / Serogroup C;
 RX MEDLINE=94018616; PubMed=8412674;
 RA Thompson S.A., Wang U.L., Sparling P.F.;
 RT "Cloning and nucleotide sequence of tlyPC, a second gene from
 RT *Neisseria meningitidis* encoding a protein similar to RTX
 RT cytotoxins.";
 RL Mol. Microbiol. 9:85-96(1993).
 CC -!- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
 CC DISEASE.
 CC -!- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED.
 CC -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
 CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
 CC ACTIVITY.
 CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, I06299; AAA99902.1; -
 DR InterPro: IPR001343; Hemlysn_Ca_bind.
 DR Pfam: PF00353; hemolysincabind; 22.
 DR PRINTS; PR00313; CABDNGRPT.
 DR PROSITE; PS00330; HEMOLYSIN_CALCITON; 17.
 KW Toxin; Calcium; Outer membrane; Repeat.
 FT DOMAIN 879 1702 43 X REPEATS, GLY-RICH.
 FT REPEAT 879 884 1.
 FT REPEAT 888 893 2.
 FT REPEAT 897 902 3.
 FT REPEAT 1016 1021 4.
 FT REPEAT 1025 1030 5.
 FT REPEAT 1034 1039 6.
 FT REPEAT 1043 1048 7.
 FT REPEAT 1052 1057 8.
 FT REPEAT 1061 1066 9.
 FT REPEAT 1070 1075 10.
 FT REPEAT 1079 1084 11.
 FT REPEAT 1088 1093 12.
 FT REPEAT 1097 1102 13.
 FT REPEAT 1216 1221 14.
 FT REPEAT 1225 1230 15.
 FT REPEAT 1234 1239 16.
 FT REPEAT 1243 1248 17.
 FT REPEAT 1252 1257 18.
 FT REPEAT 1261 1266 19.
 FT REPEAT 1270 1275 20.
 FT REPEAT 1279 1284 21.
 FT REPEAT 1288 1293 22.
 FT REPEAT 1297 1302 23.
 FT REPEAT 1416 1421 24.
 FT REPEAT 1425 1430 25.
 FT REPEAT 1434 1439 26.
 FT REPEAT 1443 1448 27.
 FT REPEAT 1452 1457 28.
 FT REPEAT 1461 1466 29.
 FT REPEAT 1470 1475 30.
 FT REPEAT 1479 1484 31.
 FT REPEAT 1488 1493 32.
 FT REPEAT 1497 1502 33.
 FT REPEAT 1616 1621 34.
 FT REPEAT 1625 1630 35.
 FT REPEAT 1634 1639 36.
 FT REPEAT 1643 1648 37.
 FT REPEAT 1652 1657 38.
 FT REPEAT 1661 1666 39.
 FT REPEAT 1670 1675 40.
 FT REPEAT 1679 1684 41.
 FT REPEAT 1688 1693 42.
 FT REPEAT 1697 1702 43.

SQ SEQUENCE 1829 AA; 197622 MW; 5C3494020A782DC8 CRC64;
 Query Match 8.6%; Score 90; DB 1; Length 1829;
 Best Local Similarity 23.2%; Req. No. 12;
 Matches 58; Conservative 33; Mismatches 69; Indels 90; Gaps 15;
 QY 1 MKIRKVMVMAMGIVISACSTPIPPKNNQLAIFQIAISLDMGK--LDOAKQO-- 56
 DB 666 MRLSTDMQTANEGIAL-----TP-----SQVADKKNALVSLDKAKAIDAARDIA 714
 QY 57 -LDAALSDROFAFAYRTIAKYQASEDA-----THQTKQRLFEKAI--ELNPK 103
 DB 715 VLDAVTGQDS-----TL--YVMSSEDAINIVKVTNDYDLAKNIYONLLFQRLQP- 765
 QY 104 DMQSYMDYGFIVQMG-----DLSGALITYFD-----KPSRAI----- 135
 DB 766 -----VLNDISFPMENDFTLPSSGLVQAFNHVKTETPPQAFVYDLAEMLAYGLR 815
 QY 136 -GYEGRVVAIENMAYIYHOYEAKSPTKDDYNNAKSALERALI---SGTOHDEIKS-- 189
 DB 816 SWYEGRRRLMAD-----YVEAKKAGKGFEDYQKVLGGETVALLAKISGTQADILLQVNG 868
 QY 190 --YDKLSDDY 197
 DB 869 FGHNNKVVSLY 878
 RESULT 19
 BSCC ECOS7 STANDARD; PRT; 1137 AA.
 ID BSCC ECOS7
 AC Q8X5M0; Q8X3J4;
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cellulose synthase operon protein C.
 GN BSCC OR Z4944/Z4945 OR ECS4410.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC *Escherichia*.
 OK NCBI_TaxId=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [3]
 RP CONCEPTUAL TRANSLATION.
 RA Michoud K.;
 RT Unpublished observations (JUN-2002).
 CC -!- FUNCTION: Required for maximal bacterial cellulose synthesis (by
 CC similarity).
 CC -!- PATHWAY: Bacterial cellulose biosynthesis.
 CC -!- SIMILARITY: BELONGS TO THE ACSC/BCSC FAMILY.
 CC -!- SIMILARITY: CONTAINS 8 TPR REPEATS.
 CC -!- CAUTION: This is a conceptual translation; one frameshift was

```

CC      introduced in position 118 to maximize the similarity with the
CC      E.coli ortholog. However, it could be a pseudogene.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE005579; AAG58672.1; ALT_FRAME.
DR      EMBL; AE005579; AAG58671.1; ALT_FRAME.
DR      EMBL; AP002565; BAB37833.1; ALT_FRAME.
DR      InterPro; IPR001440; TPR.
DR      Pfam; PF00515; TPR; 4.
DR      SMART; SM00028; TPR; 5.
KW      Cellulose biosynthesis; Repeat; TPR repeat; Complete proteome.
FT      REPEAT
FT      REPEAT 45
FT      REPEAT 249
FT      REPEAT 282
FT      REPEAT 283
FT      REPEAT 316
FT      REPEAT 331
FT      REPEAT 364
FT      REPEAT 365
FT      REPEAT 398
FT      REPEAT 441
FT      REPEAT 474
FT      REPEAT 583
FT      REPEAT 616
FT      REPEAT 690
FT      REPEAT 724
SQ      SEQUENCE 1137 AA; 125555 MW; A0F8E35AC24804B0 CRC64;

Query Match
Best Local Similarity 20.0%; Score 88.5; DB 1; Length 1137;
Matches 51; Conservative 42; Mismatches 83; Indels 79; Gaps 8;

OY      6 KMPVWAMGLVLSACSTPIPPKNNPOLAQIRTOIA-----ISLDMGKLDQAKQ 55
DB      220 KYLSIFSDGSDVNAAGS-----QLAEQOKLADPAFARAAGLAADVSGMAGKALP 270
OY      56 QLDALASDROFAPAYRTLAKVQASD-----
DB      271 ELQGVANRANPKDSEALGQAVSQGDRANAVANLEKALADPHSSNNDKNSILKYNR 330
OY      84 -----ATHQTKQRLFEKATLELNPKMOSYMDVGFVLYQMGDLSGALITYDK 130
DB      331 YWLAIQGGDAAALKANNPDRAERLFOQARNVNDTSDYAVLGLGDVAMARKDYPAAERYQQ 390
OY      131 PSRAIGYEGRVVAIENMAVYIYHOYEAAKSPTK--DDYNNAKSALERALLISGTHDEIKKS 189
DB      391 TLRM--DSGNTMAVGLANTIRQO-----SPEKAEAFIASASQRRSI-----DIERS 438
OY      190 Y--DKLSDYKLLSD 202
DB      439 LQNDRLAQQAALEN 453

RESULT 20
BCSC_ECOLI STANDARD; PRT; 1140 AA.
AC P37650; P76710;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase operon protein C.
GN BCSC OR B3530.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=362;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RT Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
```

```

RL      Nucleic Acids Res. 22:2576-2586 (1994).
RN [2]
RP REVISIONS TO 577-578.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [3]
RP CHARACTERIZATION.
RC STRAIN=ECOR10, ECOR12, and TOB1;
RX PubMed=11260463;
RA Zogaj X., Nitsch M., Rohde M., Bokranz W., Roemling U.;
RT "The multicellular morphotypes of Salmonella typhimurium and
RT Escherichia coli produce cellulose as the second component of the
RT extracellular matrix.";
RL Mol. Microbiol. 39:1452-1463 (2001).
CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis (By
CC similarity).
CC -1- PATHWAY: Bacterial cellulose biosynthesis.
CC -1- MISCELLANEOUS: The genes bcsA, bcsB, bcsZ and bcsC are
CC constitutively transcribed but cellulose synthesis occurs only
CC when adra, a putative transmembrane protein regulated by agfD, is
CC expressed. Cellulose production is abolished in E.coli K12.
CC -1- SIMILARITY: BELONGS TO THE ACSC/BCSC FAMILY.
CC -1- SIMILARITY: CONTAINS 9 TPR REPEATS.
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U00039; AAB18507.1; ALT_INIT.
DR      EMBL; AE000430; AAC76555.1; ALT_INIT.
DR      EcoGene; EG12257; bcsC.
DR      Pfam; PF00515; TPR; 3.
DR      InterPro; IPR001440; TPR.
KW      Cellulose biosynthesis; Repeat; TPR repeat; Complete proteome.
FT      REPEAT 45
FT      REPEAT 78
FT      REPEAT 95
FT      REPEAT 128
FT      REPEAT 252
FT      REPEAT 285
FT      REPEAT 286
FT      REPEAT 319
FT      REPEAT 334
FT      REPEAT 367
FT      REPEAT 401
FT      REPEAT 444
FT      REPEAT 477
FT      REPEAT 586
FT      REPEAT 619
FT      REPEAT 727
FT      REPEAT 9.
FT      CONFLICT 577
FT      CONFLICT 578
SQ      SEQUENCE 1140 AA; 125817 MW; 6811A8B8F0AF29B CRC64;

Query Match
Best Local Similarity 8.5%; Score 88.5; DB 1; Length 1140;
Matches 51; Conservative 42; Mismatches 83; Indels 79; Gaps 8;

OY      6 KMPVWAMGLVLSACSTPIPPKNNPOLAQIRTOIA-----ISLDMGKLDQAKQ 55
DB      223 KYLSIFSDGSDVNAAGS-----QLAEQOKLADPAFARAAGLAADVSGMAGKALP 273
OY      56 QLDALASDROFAPAYRTLAKVQASD-----
DB      274 ELQGVANRANPKDSEALGQAVSQGDRANAVANLEKALADPHSSNNDKNSILKYNR 333
OY      84 -----ATHQTKQRLFEKATLELNPKMOSYMDVGFVLYQMGDLSGALITYDK 130
DB      334 YWLAIQGGDAAALKANNPDRAERLFOQARNVNDTSDYAVLGLGDVAMARKDYPAAERYQQ 393
OY      131 PSRAIGYEGRVVAIENMAVYIYHOYEAAKSPTK--DDYNNAKSALERALLISGTHDEIKKS 189
```

Db 394 TIRM--DSCNTNAVGLANTYRQ-----SPEKAEFIASASQRRST-----DIIERS 441
 QY 190 Y--DKLSDYKLLSD 202
 Db 442 LQNDRLAQQAELN 456

RESULT 21
 BCC3 ACCEXY STANDARD; PRT; 1325 AA.
 AC 09W63:
 DT 15-UN-2002 (Rel. 41, Created)
 DT 15-UN-2002 (Rel. 41, Last sequence update)
 DT 15-UN-2002 (Rel. 41, Last annotation update)
 DE Cellulose synthase 1 operon protein C precursor.
 GN BCS1.
 OS Acetobacter xylinus.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Gluconacetobacter.
 OX NCBI_TaxID=28448;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCM 7664 / IFO 13693;
 RX PubMed=10382968;
 RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikenchi M.,
 Inoue Y.;
 RT "Cloning of cellulose synthase genes from Acetobacter xylinum JCM
 7664: implication of a novel set of cellulose synthase genes.";
 RL DNA Res. 6:109-115(1999).
 CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis. It
 may be involved in the formation of a membrane complex for
 extrusion of the cellulose product (by similarity).
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ACSC/BSCS FAMILY.
 CC -1- SIMILARITY: CONTAINS 8 TPR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AB015802; BAA77587.1; -;
 DR InterPro: IPR003921; CellSynth_C.
 DR InterPro: IPR001440; TPR.
 DR Pfam, PF00515; TPR; 4.
 DR PRINTS; PR01441; CELUSNTYASSEC.
 KW Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane;
 KW Signal.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1325 CELLULOSE SYNTHASE 1 OPERON PROTEIN C.
 FT REPEAT 50 83 TPR 1.
 FT REPEAT 85 117 TPR 2.
 FT REPEAT 292 325 TPR 3.
 FT REPEAT 326 359 TPR 4.
 FT REPEAT 406 439 TPR 5.
 FT REPEAT 558 591 TPR 6.
 FT REPEAT 702 735 TPR 7.
 FT REPEAT 737 765 TPR 8.
 SQ SEQUENCE 1325 AA; 142541 MW; DA4EECFADDED7AD CRC64;
 Query Match 8.4%; Score 87.5; DB 1; Length 1325;
 Best Local Similarity 25.0%; Pred. No. 14;
 Matches 34; Conservative 19; Mismatches 54; Indels 29; Gaps 5;
 QY 26 PPKNNFQAQIRFQIAISLDMGKLDQAQQLDAISADRFAPAYRTAKYQASEDT 85
 Db 284 PGGPGPDKAGLARQAGFQOQINSGRSLSAEQSPQALQINSHDASLGGMGLVSMROGDA- 342

QY 86 HQTKAORLEKAEILNPKMQSYMDYGFYLQWGDLSGALIFDKPSRAIGE--GRVVA 143
 Db 343 --AEKRRYFQEMAMADPRTADRW-----RPALA-GMEISGDYAA 378
 QY 144 IENMAYIYHQYEAK 159
 Db 379 VRQL--IAAHQYTEAK 392

RESULT 22
 PURL METAC STANDARD; PRT; 715 AA.
 AC 08RT6:
 DT 15-UN-2002 (Rel. 41, Created)
 DT 15-UN-2002 (Rel. 41, Last sequence update)
 DT 15-UN-2002 (Rel. 41, Last annotation update)
 DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
 synthase II).
 GN PURL OR MA4055.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
 Allen N., Naylor J., Stange-Thomann N., DeRellano K., Johnson R.,
 Limon L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 Zimmon A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
 Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 Metcalf W.W., Birren B.;
 RA "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity."
 RT Genome Res. 12:532-542(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide + L-
 glutamine + H(2)O = ADP + phosphate + 5'-
 phosphoribosylformylglycinamide + L-glutamate.
 CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
 CC -1- SUBUNIT: Heterodimer of two subunits, purQ and purL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AE011118; AAM07403.1; -;
 KW Purine biosynthesis; Ligase; ATP-binding; Complete proteome.
 FT NP BIND 90 101 ATP (POTENTIAL).
 SQ SEQUENCE 715 AA; 76230 MW; A3C22E07BEC6FD43 CRC64;
 Query Match 8.3%; Score 87; DB 1; Length 715;
 Best Local Similarity 27.6%; Pred. No. 7.2;
 Matches 43; Conservative 28; Mismatches 53; Indels 32; Gaps 9;
 QY 53 AKQQLDAALSDR--QFAP--AYRTLAKVYQASEDATHTQKARLEFKAEILNPKMQS 107
 Db 268 AKGGAGARILADAVQRENNMVAEITLA-----ESQERMLFEVA---DEDYDA 313
 QY 108 YMDYGFYLQWGDLSGALI-YFDK-PSAALIGEGGVAAIENMAYIYHQYEAKSPRTDD 165
 Db 314 VL-----ALVQKDLNGAVAVGVLTKETPTVEGGEIVADIPALF-----TGGAPTCEK 363

QY 166 YNNAKSALERALISGTHDEIKSYDKLSDYKLS 201
 DB 364 PSEAPTLREGKKPPEP- LKVAFLKVLSSNNIAS 398

RESULT 23

YPIA_BACSU STANDARD; PRT; 423 AA.
 ID YPIA_BACSU
 AC P54389;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DB Hypothetical protein ypia.
 GN YPIA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxId=1423;
 RN 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / Marburg;
 RX MEDLINE=96349105; PubMed=8760912;
 RA Sorokin A.V., Azavedo V., Zumstein E., Galleron N., Ehrlich S.D.,
 RA Serror P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 RT the sexA and kds loci cloned in a yeast artificial chromosome.";
 RL Microbiology 142:2005-2016(1996).
 RN 12)
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunze P., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boudier L., Brans A., Braun M., Briganti S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Etrington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hult M.F., Ilaya M., Jones L.,
 RA Joriba K., Karmata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Kroch S., Kumano M.,
 RA Kurita K., Lapins A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilawa A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priesen E., Fujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccoti E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpeira P., Tognoni A.,
 RA Toasato V., Uchiyama S., Vandenbol M., Vannier P., Vassaretto A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Wellenreger T.,
 RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RA "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RT Nature 390:249-256(1997).
 RL Nature 390:249-256(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL, L47709; AAB38432.1; -
 DR EMBL, Z99115; CAB14175.1; -

DR Subtilist; BG11496; ypia.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 5.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 423 AA; 48321 MW; 00887D063282DD7A CRC64;

Query Match 8.3%; Score 86.5; DB 1; Length 423;
 Best Local Similarity 24.6%; Pred. No. 4.2;
 Matches 31; Conservative 16; Mismatches 34; Indels 45; Gaps 4;

QY 45 LDMKLDQAKQDLALSDAPAFAPAYRTAKYQASDATHQYKQRLFEKALFLNPKD 104
 DB 283 LKIGSEBEGKTLDEGALLDGFVEALHTLLAVHKEED-----YQIIDL-IOE 331
 QY 105 MGSYMD-----YGFVYQNGDLSGALIVDPKSRAGYGRVVAIEMAYIYHQAFAKS 160
 DB 332 VRSYGEEDPKYNNWYLAS-----AYTELOYEAKQ 361

QY 161 PTKDYL 166
 DB 362 SPEAY 367

RESULT 24

IFIT1_MOUSE STANDARD; PRT; 463 AA.

ID IFIT1_MOUSE
 AC O64282;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Interferon-induced protein with tetratricopeptide repeats 1 (IFIT-1)
 DE (Interferon-induced 56 kDa protein) (IFI-56k) (Glucocorticoid-
 DE attenuated response gene 16 protein) (GARG-16).
 GN IFIT1 OR IFI56 OR GARG16.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 CX NCBI_TaxId=10090;
 RN 11)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95203869; PubMed=7896268;
 RA Bluyssen H.A., Vilestra R.J., Faber P.W., Smit E.M., Hagemeljer A.,
 RA Trippan J.;
 RT "Structure, chromosome localization, and regulation of expression of
 RT the interferon-regulated mouse Ifi54/Ifi56 gene family.";
 RL Genomics 24:137-148(1994).
 RN 12)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96239145; PubMed=8660659;
 RA Smith J.B., Herschman H.R.;
 RT "The glucocorticoid attenuated response genes GARG-16, GARG-39, and
 RT tetratricopeptide repeat domains.";
 RT Arch. Biochem. Biophys. 330:290-300(1996).
 RT Arch. Biochem. Biophys. 330:290-300(1996).
 CC -1- SIMILARITY: BELONGS TO THE IFIT FAMILY.
 CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL, S77715; AAB33631.2; -
 CC EMBL, S77714; AAB33631.2; JOINED.
 DR EMBL, U43084; AAC52625.1; -
 DR MGD, MGI:99450; Ifit1.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 6.
 DR Repeat; TPR repeat; Interferon induction.
 FT REPEAT 52 85 TPR 1.

RA Hortonuchi S., Sone Y., Mori H., Sakai F., Hayashi T.;
 RT "Control of expression by the cellulose synthase (bcsa) promoter
 RL region from *Acetobacter xylinum* BPR 2001.";
 CC Gene 213:93-100(1998).
 CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis. It
 CC may be involved in the formation of a membrane complex for
 CC extrusion of the cellulose product (By similarity).
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -1- SIMILARITY: BELONGS TO THE AGSC/BCSC FAMILY.
 CC -1- SIMILARITY: CONTAINS 6 TPR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AB010645; BAA31465.1; -;
 CC InterPro; IPR003921; CellSynth_C.
 CC DR InterPro; IPR001440; TPR.
 CC DR Pfam; PF00515; TPR; 3.
 CC DR PRINTS; PR01441; CELLSYNTHASEC.
 CC KW Cellulose biosynthesis; Repeat; TPR repeat; Membrane;
 CC Outer membrane; Signal.
 CC FT SIGNAL 1 30 POTENTIAL.
 CC FT CHAIN 31 1326 CELLULOSE SYNTHASE OPERON PROTEIN C.
 CC FT REPEAT 49 82 TPR 1.
 CC FT REPEAT 291 324 TPR 2.
 CC FT REPEAT 325 358 TPR 3.
 CC FT REPEAT 405 438 TPR 4.
 CC FT REPEAT 557 590 TPR 5.
 CC FT REPEAT 701 734 TPR 6.
 CC FT REPEAT 1326 AA; 142103 MW; FCA42350C46C4A9 CRC64;
 CC SQ
 CC -----
 CC Query Match 8.2%; Score 85.5; DB 1; Length 1326;
 CC Best Local Similarity 23.7%; Pred. NO. 20;
 CC Matches 32; Conservative 18; Mismatches 56; Indels 27; Gaps 4;
 CC
 CC QY 26 PPKNNPQIAQIRTOIAISLDLMDGKLDQAKQOALDASDRPAPARYRTIARVQAASEDAT 85
 CC DB 283 PPGGPPDAGLAGLARGOYQOOLNAGRLSAEQSFQSLQINSHDADSLGGMGVLSMKGQDT 341
 CC QY 86 HQTKAQRLEFKALINPKDMOSYNDGYLVQMGDLSGALLYFDKPSQA-IGYERVVAI 144
 CC DB 342 --AEKHRYFEEMADMDPTADRW-----RPALAGMAVSGDYAAV 378
 CC QY 145 ENMAIYVHYOYEAAK 159
 CC DB 379 ROL--IAHQYTEAK 391
 CC
 CC RESULT 27
 CC PDRC_YEAST STANDARD; PRT; 1511 AA.
 CC ID Q02785;
 CC AC Q02785;
 CC DT 15-JUL-1999 (Rel. 38, Created)
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE ATP-dependent permease PDR12.
 CC GN PDR12 OR YPL058C OR LPR14C.
 CC OS *Saccharomyces cerevisiae* (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; *Saccharomycotina*; *Saccharomycetes*;
 CC OC *Saccharomycetales*; *Saccharomycetaceae*; *Saccharomycetes*.
 CC OX NCBI_TaxID=4932;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=8286C / AB972;
 CC RX MEDLINE=97313371; PubMed=9169875;
 CC RA Busey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgore W.,
 CC Araujo R., Aparicio A., Bartell B.G., Badcock K., Benes V.,

RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E.S., Churcher C.M., Cosser F., Davis K., Davis R.W.,
 RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duetschhoef A.,
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hedling U., Heumann K., Hilbert H., Hillier L.,
 RA Hunnicke-Smith S., Hyman R., Johnston M., Kaiman S., Kleene K.,
 RA Komp C., Kurit O., Laskhari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marathe R., Messenguy F., Mewes H.-W., Miripati S., Moestl D.,
 RA Mueller-Auer S., Namath A., Newwisch U., Oefner P., Pearson D.,
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
 RA Uristarazu U.A., Ushinsky S., Vienderleis F., Vissers S., Voss H.,
 RA Walsh S.V., Wambut R., Wang Y., Wedler E., Wedler H., Winnett E.,
 RA Zhong W.W., Zollner A., Vo D.H., Hand J.;
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI.";
 CC Nature 387:103-105(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. PDR5 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; U39205; AAB68307.1; -;
 CC DR SGD; S0005979; PDR12.
 CC DR InterPro; IPR003593; AAA ATPase.
 CC DR InterPro; IPR003439; ABC transporter.
 CC DR Pfam; PF00005; ABC tran; 2.
 CC DR ProDom; PD000006; ABC transporter; 2.
 CC DR SMART; SM00382; AAA; 1.
 CC DR PROSITE; PS00211; ABC TRANSPORTER; FALSE_NEG.
 CC KW ATP-binding; Transmembrane; Cytoplasmic (Potential).
 CC FT DOMAIN 1 508
 CC FT TRANSMEM 509 529 POTENTIAL.
 CC FT TRANSMEM 549 569 POTENTIAL.
 CC FT TRANSMEM 598 618 POTENTIAL.
 CC FT TRANSMEM 623 643 POTENTIAL.
 CC FT TRANSMEM 658 678 POTENTIAL.
 CC FT TRANSMEM 766 786 POTENTIAL.
 CC FT DOMAIN 787 1182 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 1183 1203 POTENTIAL.
 CC FT TRANSMEM 1205 1225 POTENTIAL.
 CC FT TRANSMEM 1255 1275 POTENTIAL.
 CC FT TRANSMEM 1292 1312 POTENTIAL.
 CC FT TRANSMEM 1319 1339 POTENTIAL.
 CC FT TRANSMEM 1445 1465 POTENTIAL.
 CC FT DOMAIN 1466 1511 CYTOPLASMIC (POTENTIAL).
 CC FT NP_BIND 878 885 ATP (POTENTIAL).
 CC FT CARBOHYD 1405 1405 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT SEQUENCE 1511 AA; 171064 MW; 4962762AAE1997FC CRC64;
 CC SQ
 CC -----
 CC Query Match 8.1%; Score 85; DB 1; Length 1511;
 CC Best Local Similarity 21.3%; Pred. NO. 26;
 CC Matches 37; Conservative 25; Mismatches 48; Indels 64; Gaps 7;
 CC
 CC QY 51 DOAKQDPAALISAROPAPART-----LAKVYQASDATHQTKAQLFEKALELN 101
 CC DB 320 DNATRGDA--STLFEPAQARTATNMVNSAIYAIYQAGNN-----IYLFPKTYVL 370
 CC QY 102 PKDMOSTYMDGYFVLVQMGDLSGALLYFDKPSRAIGRGVVAIENMAIYVHYOYEAAKSP 161
 CC DB 371 -----YNGROIYFGPADKAVGY-----FGPMGVKXENRMTSAEFL 405
 CC QY 162 T-----KDDYNAKSALEBALISGTQHDIKRSYKLSLDYK 198
 CC DB 406 TSVTVDFENRTLDIKPGYEDKVPKSSSEFBEYTWLNSDYLRLTYD---DYQ 455
 CC
 CC RESULT 28

```

IEFS_HUMAN
ID IEFS_HUMAN STANDARD; PRT; 543 AA.
AC P31948;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Stress-induced-phosphoprotein 1 (SRI1) (Hsp70/Hsp90-organizing
protein) (transformation-sensitive protein IEF SSP 3521).
GN SRI1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92235077; PubMed=1569099;
RA Honore B., Leftere H., Madsen P., Rasmussen H.H., Vandekerckhove J.,
RA Celis J.E.;
RT "Molecular cloning and expression of a transformation-sensitive human
protein containing the TPR motif and sharing identity to the stress-
inducible yeast protein SRI1."
RL J. Biol. Chem. 267:8485-8491(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 101-109; 352-364 AND 374-381.
RC TISSUE=Keratinocytes;
RX MEDLINE=93162043; PubMed=1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
protein database of normal human epidermal keratinocytes."
RL Electrophoresis 13:960-969(1992).
RN [4]
RP -1- SIMILARITY: TO YEAST SRI1.
CC -1- SIMILARITY: CONTAINS 9 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M86752; AAA58682.1; -
CC DR EMBL; BC002987; AAH02987.1; -
CC DR PIR; A38093; A38093.
CC DR Aarhus/Chent-2DPAGE; 2410; IEF.
CC DR Genew; HGNC:11387; STIPL.
CC DR MIM; 605063; -
CC DR InterPro; IPR001440; TPR.
CC DR Pfam; PF00515; TPR; 9.
CC DR SMART; SM00026; TPR; 9.
CC KW TPR repeat; Repeat.
FT REPEAT 4
FT REPEAT 37
FT REPEAT 39
FT REPEAT 71
FT REPEAT 73
FT REPEAT 105
FT REPEAT 225
FT REPEAT 258
FT REPEAT 260
FT REPEAT 292
FT REPEAT 300
FT REPEAT 333
FT REPEAT 360
FT REPEAT 393
FT REPEAT 427
FT REPEAT 438
FT REPEAT 461
FT REPEAT 490
FT DOMAIN 543
FT SEQUENCE 543 AA; 62639 MW; 8B58ECA13825CE0B CRC64;
Query Match 8.1%; Score 84.5; DB 1; Length 543;
Best Local Similarity 22.7%; Pred. No. 8.4;
Matches 44; Conservative 30; Mismatches 67; Indels 53; Gaps 9;

```

```

QY 28 KNKP-----QLAQIRQLAISL-LPWGKLDOAKQDLAALSDROPAPAPVRLAKYVQ 79
DB 160 RKNPBDLCTKQLDPRIMTTLISVLGVDIGSNDDEEE-----IATPPPPPKETKTEP 213
QY 80 ASEDATHTQTKAQRLEFKAIEINPKDMOSYMDYGFYVQMDLSGALLYFDR-----PSRA 134
DB 214 MEEDLP-ENKQALKEK-----ELGNDAYKKKQFDPVALKHYDAKEIDPT-- 257
QY 135 IGYEGRVAIENMAYIYHGYEAKSPKPDYNNAKSLERALLSGTGHDE----IKKSY 190
DB 258 -----NMITYNQAAVYE-----KGDYKRCRELCEKAIEVGRNEDYRQIATAY 303
QY 191 DKLSDDYKLSDDYK 204
DB 304 ARIGNSYKREKRYK 317

RESULT 29
OM70_YEAST STANDARD; PRT; 617 AA.
AC P07213;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial precursor proteins import receptor (70 kDa mitochondrial
outer membrane protein) (Translocase of outer membrane TOM70).
GN TOM70 OR MAST70 OR OMP1 OR YNL121C OR N1905.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84131928; PubMed=6365533;
RA Hase T., Riezman H., Suda K., Schatz G.;
RT "Import of proteins into mitochondria: nucleotide sequence of the
RT gene for a 70-kd protein of the yeast mitochondrial outer membrane."
RL EMBO J. 2:2169-2172(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97245296; PubMed=9090055;
RA de Antoni A., D'Angelo M., Dal Pero F., Sartorello F., Pandolfo D.,
RA Pallavicini A., Lanfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-13p from chromosome XIV of
RT Saccharomyces cerevisiae reveals an unusually high number of
RT overlapping open reading frames."
RL Yeast 13:261-266(1997).
RN [3]
RP TPR REPEATS.
RX MEDLINE=90309987; PubMed=1694969;
RA Boguski M.S., Sikorski R.S., Hieter P.A., Goebel M.;
RT "Expanding family."
RL Nature 346:114-114(1990).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=91006012; PubMed=2170106;
RA Hines V., Brandt A., Griffiths G., Horstmann H., Bruesch H.,
RA Schatz G.;
RT "Protein import into yeast mitochondria is accelerated by the outer
RT membrane protein MAST70."
RL EMBO J. 9:3191-3200(1990).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=93107043; PubMed=8380165;
RA Hines V., Schatz G.;
RT "Precursor binding to yeast mitochondria. A general role for the
RT outer membrane protein MAST70."
RL J. Biol. Chem. 268:449-454(1993).
CC -1- FUNCTION: RECEPTOR THAT ACCELERATES THE IMPORT OF ALL
CC MITOCHONDRIAL PRECURSOR PROTEINS. THOUGHT TO ACT IN CONJUNCTION
CC WITH MAST37.
CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC LEAST 8 DIFFERENT PROTEINS (TOM5, TOM6, TOM7, TOM20, TOM22,

```

CC TOM37, TOM40 AND TOM70). FORMS A HETERODIMER WITH MAS37.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein; Mitochondrial
 CC outer membrane.
 CC -1- SIMILARITY: CONTAINS 9 TPR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, X05585; CAA29085.1; -;
 CC DR EMBL, Z69382; CAA93386.1; -;
 CC DR EMBL, Z71397; CAA96002.1; -;
 CC PIR, S05884; MMEYO.
 CC HSRP, P53041; IAI7.
 CC SGD, S0005065; TOM70.
 CC InterPro: IPR001440; TPR.
 CC Pfam: PF00515; TPR; 8.
 CC SMART: SM00028; TPR; 4.
 CC TIGRfam: TIGR00590; 3a801809; 1.
 CC DR Microchondrion; Outer membrane; Transmembrane; Repeat; TPR repeat.
 CC KM DOMAIN 1 10 INTERMEMBRANE (POTENTIAL).
 CC FT TRANSMEM 11 30 POTENTIAL.
 CC FT DOMAIN 31 617 CYTOPLASMIC (POTENTIAL).
 CC FT REPEAT 99 132 TPR 1.
 CC FT REPEAT 134 165 TPR 2.
 CC FT REPEAT 281 315 TPR 3.
 CC FT REPEAT 363 396 TPR 4.
 CC FT REPEAT 397 430 TPR 5.
 CC FT REPEAT 432 464 TPR 6.
 CC FT REPEAT 465 498 TPR 7.
 CC FT REPEAT 505 541 TPR 8.
 CC FT REPEAT 542 575 TPR 9.
 CC FT CONFLICT 185 185 A -> R (IN REF. 1).
 CC FT SEQUENCE 617 AA; 70123 MW; 664BFC32CCE39A2 CRC64;
 CC SQ
 CC
 CC Query Match 8.1%; Score 84; DB 1; Length 617;
 CC Best Local Similarity 24.2%; Pred. No. 11;
 CC Matches 37; Conservative 25; Mismatches 69; Indels 22; Gaps 5;
 CC
 CC QY 58 DAALSDADRFAPAYRTLAKVYQASEDATHTQ--TKAQLFEKALIELNPKMOSYMDGYFL 115
 CC DB 274 DESNADKELNGLSNLYKRSPESTYDKADESTYKARLFEEEDJDNEDKELKELATSL 333
 CC QY 116 VQWG-----DLGALIFDPKSPRAIGEGRVVAIENMAYIYHQYEAKSP 161
 CC DB 334 EHTGIFKFLKNDPLGAHEDIKAIELFPVNSYI-YMALIMADRDSTREYNYFDKA--- 389
 CC QY 162 TKDDVNNAKSALERALLISG--TQHDEIKSKYDK 192
 CC DB 390 LKLDSSNNSYYHRCQNNFLLQNYDQAGDPDK 422
 CC
 CC RESULT 30
 CC Y751_SYNY3 STANDARD; PRT; 248 AA.
 CC AC Q04737;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC GN Hypothetical protein elr0751.
 CC OS SRR0751.
 CC OS Synecocystis sp. (strain PCC 6803).
 CC OC Bacteria; Cyanobacteria; Chroococcales; Synecocystes.
 CC OX NCBI_TaxID=1148;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=92330029; PubMed=1368342;
 CC RA Ogura Y., Takemura M., Oda K., Ohca E., Yamato K., Fukuzawa H.,
 CC Ohyama K.;

RT "Cloning and nucleotide sequence of a fixC-ORF469 gene cluster of
 RT Synecocystis PCC6803: conservation with liverwort chloroplast fixC-
 RT ORF465 and nit operon";
 RT Biosci. Biotechnol. Biochem. 56:788-793(1992).
 RT [2]
 RT RP SEQUENCE FROM N.A.
 RT RX MEDLINE=97061201; PubMed=8905231;
 RT Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RT Miyajima N., Hironaka M., Sugitani M., Sasamoto S., Kimura T.,
 RT Honouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
 RT Shimpou S., Takeuchi C., Wada T., Matenabe A., Yamada M.,
 RT Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions";
 RT DNA Res. 3:109-136(1996).
 CC -1- SIMILARITY: CONTAINS 4 TPR REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC CC EMBL, D10474; BAA01277.1; -;
 CC DR EMBL, D90817; BAA18748.1; -;
 CC DR InterPro: IPR001440; TPR.
 CC DR Pfam: PF00515; TPR; 5.
 CC DR SMART: SM00028; TPR; 4.
 CC KM Hypothetical protein; Repeat; TPR repeat; Complete proteome.
 CC FT REPEAT 61 94 TPR 1.
 CC FT REPEAT 95 128 TPR 2.
 CC FT REPEAT 129 162 TPR 3.
 CC FT REPEAT 163 196 TPR 4.
 CC FT SEQUENCE 248 AA; 27753 MW; AB2DC380BF6A4A7 CRC64;
 CC SQ
 CC
 CC Query Match 8.0%; Score 83.5; DB 1; Length 248;
 CC Best Local Similarity 25.0%; Pred. No. 4;
 CC Matches 21; Conservative 19; Mismatches 41; Indels 3; Gaps 2;
 CC
 CC QY 47 MGKLDQAKQDUALSDROPAPAYRTLAKVYQASEDATHTQKARLFEKALIELNPKMQ 106
 CC DB 108 LGNDQALADYGRSLADRYIIPYIRGNLYSQGD--HTHTALQD-FTQALITDPNRKX 164
 CC QY 107 SYMDYGYLVQWGLSGALIFDK 130
 CC DB 165 AYVNRANSYFQDGGYQAQIADYNR 188
 CC
 CC RESULT 31
 CC 6P22_YEAST STANDARD; PRT; 397 AA.
 CC AC Q12471;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 15-JUL-1989 (Rel. 38, Last annotation update)
 CC DE 6-phosphofructo-2-kinase 2 (EC 2.7.1.105) (Phosphofructokinase 2 II)
 CC (6PF-2-K 2).
 CC GN 6PF27 OR YOL136C.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC OX NCBI_TaxID=4932;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=WA-1A;
 CC RX MEDLINE=97014370; PubMed=8861205;
 CC RA Boles E., Goehmann H.W.H., Zimmermann F.K.;
 CC "Cloning of a second gene encoding 5-phosphofructo-2-kinase in yeast,
 CC RT and characterization of mutant strains without fructose-2,6-
 CC RT bisphosphate.";

```

RL Mol. Microbiol. 20:65-76(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97051593; PubMed=8896270;
RA Aldea M., Piedrafito L., Casas C., Casamayor A., Khalid H.,
RA Balcells L., Arino J., Herrero E.,
RT "Sequence analysis of a 12 801 bp fragment of the left arm of yeast
RT chromosome XV containing a putative 6-phosphofructo-2-kinase gene, a
RT six other open reading frames";
RL Yeast 12:1053-1058(1996).
CC -1- FUNCTION: SYNTHESIS OF FRUCTOSE 2,6-BISPHOSPHATE.
CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 2,6-bisphosphate.
CC -1- INDUCTION: BY GLUCOSE AND FRUCTOSE, BUT NOT BY GALACTOSE OR
CC MALTOSE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X90861; CAA6371.1; -
CC EMBL; Z74878; CAA9157.1; -
CC EMBL; X95465; CAA6473.1; -
CC SGD; S0005496; PFK27.
CC InterPro: IPR000546; 6PF2K.
CC Pfam; PF01591; 6PF2K; 1.
CC Transferase; Kinase; ATP-binding.
CC NP BIND 103 110
CC FT DOMAIN 85 305 6-PHOSPHOFRUCTO-2-KINASE.
CC FT ACT SITE 197 197 POTENTIAL.
CC FT ACT SITE 235 235 POTENTIAL.
CC FT BINDING 269 269 TO FRU-6-P (BY SIMILARITY).
CC FT SEQUENCE 397 AA; 45317 MW; F8A36BBE135A55C CRC64;
SQ
Query Match
Best Local Similarity 8.0%; Score 83.5; DB 1; Length 397;
Matches 45; Conservative 32; Mismatches 70; Indels 49; Gaps 9;
QY 28 KNNP-----QLAQRTQIAISLLDMGKLDQAKQDLDAISADROPAPAYRTIA 75
DB 121 KNNPLTNSLRCKVFNAGKIRROISCATISKPLLSNTSSSEDLFNPNNDKETVARIITQ 180
QY 76 KYVQASE-----DATHQTKAOR--LFEKAIELNPKDMQSY-----MDYGF 113
DB 181 KLFHEINNECDVGIDATNNTERRRFFEEVCSNTDELSEFNLPVPIILQVSCNRSP 240
QY 114 YLVQMGLSGALTYPRKPSRAIGYEGRAVAIENMAYIYHQYEAKSPTKDDYNNKSA 173
DB 241 IKVHNKSPNEEDYLPK-----YE--LAKDFAKRLKHYVQGFPSLDEFNQ-----I 288
QY 174 ERALLISGTQDEIKKS 189
DB 289 HRYI---SQHEIDTS 301
RESULT 32
BCC1 ACCEXY STANDARD; PRT; 1319 AA.
AC P19450;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase operon protein C precursor.
GN BCSC.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.

```

```

OX NCBI_TaxID=26448;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1306-3;
RX MEDLINE=91045951; PubMed=2146681;
RA Wong H.C., Fear A.L., Calhoun R.D., Eichinger G.H., Mayer R.,
RA Amikam D., Benzman M., Gelfand D.H., Meade J.H., Emerick A.W.,
RA Bruner R., Ben-Bassat A., Tal R.;
RT "Genetic organization of the cellulose synthase operon in Acetobacter
RT xylinum";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8130-8134(1990).
CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis. It
CC may be involved in the formation of a membrane complex for
CC extrusion of the cellulose product.
CC -1- PATHWAY: Bacterial cellulose biosynthesis.
CC -1- SUBCELLULAR LOCATION: Outer membrane (potential).
CC -1- SIMILARITY: BELONGS TO THE ACS/BSC FAMILY.
CC -1- SIMILARITY: CONTAINS 9 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M37202; AAA21866.1; -
CC PIR; C43735; C43735.
CC InterPro: IPR003921; Cellsynth_C.
CC InterPro: IPR001440; TPR.
CC Pfam; PF00515; TPR; 3.
CC DR PRINTS; PR01441; CELLSTHASEC.
CC Cellulose biosynthetis; Repeat; TPR repeat; Membrane; Outer membrane;
CC Signal.
CC KW SIGNAL 1 46
CC FT CHAIN 47 1319 CELLULOSE SYNTHASE OPERON PROTEIN C.
CC FT REPEAT 49 82 TPR 1.
CC FT REPEAT 84 116 TPR 2.
CC FT REPEAT 281 324 TPR 3.
CC FT REPEAT 325 358 TPR 4.
CC FT REPEAT 405 438 TPR 5.
CC FT REPEAT 557 590 TPR 6.
CC FT REPEAT 701 734 TPR 7.
CC FT REPEAT 736 768 TPR 8.
CC FT REPEAT 1059 1094 TPR 9.
CC FT SEQUENCE 1319 AA; 141431 MW; 48C6BEAF6C59A7F7 CRC64;
SQ
Query Match
Best Local Similarity 8.0%; Score 83.5; DB 1; Length 1319;
Matches 31; Conservative 20; Mismatches 57; Indels 27; Gaps 4;
QY 26 PPKNNPQLAQRTQIAISLLDMGKLDQAKQDLDAISADROPAPAYRTIAKYQASEPAT 85
DB 283 PPGGPDDVAGLRAQGYOQLNAGRLAAEQSFQSLQINSHDADSLGGGLVSMRQGT- 341
QY 86 HQTKAQLFEKAIELNPKDMQSYMDYGYLVQMGLSGALTYFDRSPRA-IGYEGRAVAI 144
DB 342 --AEARRYFEAMADPTADRW-----RPLANGMAVSEGYASV 378
QY 145 ENMAYIYHQYEAK 159
DB 379 RQL--IAAHQYTEAK 391
RESULT 33
OM70 HUMAN STANDARD; PRT; 608 AA.
AC O94826;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mitochondrial precursor proteins import receptor (Translocase of outer

```

```

DE membrane TOM70).
GN TOM70A OR KIA0719.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Ref. 5:277-286(1998).
CC -1- FUNCTION: RECEPTOR THAT ACCELERATES THE IMPORT OF ALL
CC MITOCHONDRIAL PRECURSOR PROTEINS (BY SIMILARITY).
CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC LEAST 8 DIFFERENT PROTEINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC outer membrane (By similarity).
CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB018262; BAA34439.1; -
DR HSSP; P53041; 1A17
DR GeneW; HGNC:11985; TOM70A.
DR MIM; 606081; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 9.
DR SMART; SM00028; TPR; 8.
KW Mitochondrion; Outer membrane; Transmembrane; Repeat; TPR repeat.
FT DOMAIN 1 8 INTERMEMBRANE (POTENTIAL).
FT TRANSMEM 9 29 POTENTIAL.
FT DOMAIN 30 608 CYTOPLASMIC (POTENTIAL).
FT REPEAT 114 147 TPR 1.
FT REPEAT 153 186 TPR 2.
FT REPEAT 194 294 TPR 3.
FT REPEAT 329 362 TPR 4.
FT REPEAT 367 400 TPR 5.
FT REPEAT 401 434 TPR 6.
FT REPEAT 440 475 TPR 7.
FT REPEAT 476 509 TPR 8.
FT REPEAT 511 544 TPR 9.
FT REPEAT 545 578 TPR 10.
SQ SEQUENCE 608 AA; 67454 MW; 5A8F5CAAA8582480 CRC64;

Query Match 8.0%; Score 83; DB 1; Length 608;
Best Local Similarity 24.3%; Pred. No. 13;
Matches 42; Conservative 25; Mismatches 90; Indels 16; Gaps 6;

OY 31 POLAIFRTQIAISLIDMGLKDLDAQKQDLDAISADRFAPAYRTAKV-----QASEDATH 86
DB 432 PESALAAQAKCFALYRQAVYTGNNSSQIOAMKGFEEVKKFRCAGYALYAQALTDDQQ 491
OY 87 QTRKQRLPEKALIELNPKMOSYMDGFYLVQ-MGLSGALYFDPKSAIGEGV-VAI 144
DB 492 FGKADMDTKDCLDEPDNATYVHGKLLQQLQWKQDLDRGL--ELISFAIELDNKCDPAY 548
OY 145 ENMAIYVHQ--YEAAKSPTKDDYNNNAKSALEBALI-----SGTHDEIKSY 190
DB 549 ETMGITIEVRGNGMEKALIDMFKAILAKSEMEWMLYLGLCAAHNQTEVAKKY 601

RESULT 34

```

```

OM70 MOUSE
ID OM70 MOUSE STANDARD; PRT; 611 AA.
AC 09CZM5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial precursor proteins import receptor (Translocase of outer
DE membrane TOM70).
GN TOM70A OR D16WSU109E.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kanakawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Flisbachmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guentlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Matchom L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nozode P., Ring B., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Saeki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Sasaki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontuski S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
CC -1- FUNCTION: RECEPTOR THAT ACCELERATES THE IMPORT OF ALL
CC MITOCHONDRIAL PRECURSOR PROTEINS (BY SIMILARITY).
CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC LEAST 8 DIFFERENT PROTEINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC outer membrane (By similarity).
CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AK012084; BAB28018.1; -
DR HSSP; P53041; 1A17.
DR MGD; MGI:106295; D16Wsu109e.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 10.
DR SMART; SM00028; TPR; 10.
KW Mitochondrion; Outer membrane; Transmembrane; Repeat; TPR repeat.
FT DOMAIN 1 9 INTERMEMBRANE (POTENTIAL).
FT TRANSMEM 10 30 POTENTIAL.
FT DOMAIN 31 611 CYTOPLASMIC (POTENTIAL).
FT REPEAT 117 150 TPR 1.
FT REPEAT 156 189 TPR 2.
FT REPEAT 197 297 TPR 3.
FT REPEAT 332 365 TPR 4.
FT REPEAT 370 403 TPR 5.
FT REPEAT 404 437 TPR 6.
FT REPEAT 445 478 TPR 7.
FT REPEAT 479 512 TPR 8.
FT REPEAT 514 547 TPR 9.
FT REPEAT 548 581 TPR 10.

```


RL FEBS Lett. 427:79-84(1998).

CC -1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING

CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. ITS MOTOR

CC ACTIVITY IS DIRECTED TOWARD THE MICROTUBULE'S PLUS END. THE SPEED

CC OF THIS MOTOR IS 4-5 TIMES FASTER THAN ITS ANIMAL COUNTERPARTS.

CC -1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-

CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF

CC KINESIN (IT HYDROLYSES ATP AND BINDS MICROTUBULE), A CENTRAL

CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN

CC DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH

CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),

CC VESICLES AND MEMBRANOUS ORGANELLES.

CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AJ225894; CAA12647.1; -.

CC DR HSSP; P33176; 1B92.

CC DR InterPro; IPR001752; kinesin_motor.

CC DR Pfam; PF00225; kinesin; 1.

CC DR PRINTS; PR00380; KINESINHEAVY.

CC DR SMART; SM00129; KISC; 1.

CC DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.

CC DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.

CC KM Motor protein; Microtubules; ATP-binding; Coiled coil.

CC FT DOMAIN 1 341 KINESIN-MOTOR (BY SIMILARITY).

CC FT DOMAIN 342 887 COILED COIL (POTENTIAL).

CC FT NP_BIND 888 934 GLOBULAR.

CC FT NP_BIND 87 94 ATP (POTENTIAL).

CC FT NP_BIND 237 244 ATP (POTENTIAL).

CC SQ SEQUENCE 935 AA; 105040 MW; CFBDF33827E28 CRC64;

Query Match 7.9%; Score 82.5; DB 1; Length 935;

Best Local Similarity 20.1%; Pred. No. 24; Matches 49; Conservative 34; Mismatches 74; Indels 87; Gaps 8;

QY 28 KNNPOLAQRITQIAISL-----LDMGKLPQAQKQDPAISADROFAPAVRTIAK 76

DB 473 KENQMTSELRLRLQLOKVSYESKENATVUSLKEANDLMAELBEUK-----KATSE 526

QY 77 VYQASEDAHTQKQRLPEKALE-----LNPKQWQ----- 106

DB 527 MRQAHKDATSDSKERKAEMKQNMMSGFDPGSLINDKERQINNALSKLDGEQQTLYVED 586

QY 107 -----SYMDGFIYVQMGDISGALITPDK--PSRAIGYGRVAIR----- 145

DB 587 LVSLRRLAESQMLVEQHTKTIISDLA-----DKDAMEKKYELGRLGALKEKEYEELLD 641

QY 146 -----NMAYIYHYQEAASPTKDYNNAKSALERALISGTQHDDEIKSYDKLSPY 197

DB 642 KTIABEAMNOMADVDNISALTKLEAQAQAEKKEVQAKI-----DLKRLDKQSGH 695

QY 198 KLS 201

DB 696 EKLS 699

RESULT 37

DPOM_MOUSE STANDARD; PRT; 496 AA.

AC Q9J1M4; Q9J1M9;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE DNA polymerase mu (EC 2.7.7.7) (Pol Mu).

GN POLM.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxId=10090;

RP SEQUENCE FROM N.A.

RX MEDLINE;20440416; PubMed=10982892;

RA Aoufouchi S., Flatter E., Dahan A., Falli A., Bertocci B., Storck S.,

RA Delbos F., Cocca L., Gupta N., Weill J.-C., Reynaud C.-A.;

RT "Two novel human and mouse DNA polymerases of the polk family.";

RL Nucleic Acids Res. 28:3684-3693(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CD-1;

RA Blanco L., Bernad A.;

RT "DNA polymerase mu (Pol mu), homologous to Tdt, could act as a DNA

RT mutator in eukaryotic cells.";

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Seems to act as an Ig mutase which is responsible for

CC immunoglobulin (Ig) gene hypermutation.

CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate

CC + [DNA] (N).

CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-X FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF176098; AAF27552.1; -.

CC DR EMBL; AJ251804; CAB71154.1; -.

CC DR MGD; MGI:1860191; Polm.

CC DR InterPro; IPR001357; BRCT.

CC DR InterPro; IPR002054; DNA_polX.

CC DR Pfam; PF00533; BRCT; 1.

CC DR PRINTS; PR00869; DNAPOLX.

CC DR SMART; SM00292; BRCT; 1.

CC DR SMART; SM00483; POLX; 1.

CC DR PROSITE; PS00172; BRCT; 1.

CC DR PROSITE; PS00522; DNA POLYMERASE X; 1.

CC KM Transferase; Nucleotidyltransferase.

CC FT DOMAIN 23 122 BRCT.

CC FT CONFLICT 18 18 D -> A (IN REF. 2).

CC FT CONFLICT 386 386 A -> V (IN REF. 2).

CC SQ SEQUENCE 496 AA; 55534 MW; 7C73AC22C40D3C5D CRC64;

Query Match 7.9%; Score 82; DB 1; Length 496;

Best Local Similarity 20.0%; Pred. No. 12; Matches 44; Conservative 37; Mismatches 81; Indels 58; Gaps 8;

QY 10 VMAMGLVLSACSTPIPPKNNPOL-----AQIRQLAISLDMGKLPQAQKQDA----- 59

DB 176 LLSFSRAASVLSKSPCPASISQLHGUPYFBEHSTRVYQELLEHGTCGEVQKCSERYQ 235

QY 60 -----ALSADROFAPAVRTIAKYVQASEDAHTQTKA----- 90

DB 236 TMKLFQVGVGVGTANRWYQEGRLDELAEQORLTOQOKAGQAYVODLSTPRADA 295

QY 91 ---QRLPEKALELNPKQMOSIMDYGFIVYVQMGDISG---ALIFYDKSRALIGRVA 143

DB 296 EALQQLLEAAVRQTLPGAIVTLTGAF---RRGKLQGHVDVPLTHPBGQEVGLLPKWS 352

QY 144 -IENNAVYHYQEAASPTKDYNNAK-----SALERA 176

DB 353 CLOSGQGLVTHQYH--RSHLADSAHNLKQKRSSTYMDAFERS 390

RESULT 38

YAG7_PSEAE

```

ID YAG7_PSEAE STANDARD; PRT; 590 AA.
AC P42810; Q9HVC6;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein P4667.
GN P4667.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=95189718; PubMed=7883699;
RA Hungerer C., Troup B., Roemling U., Jahn D.;
RT "Regulation of the hema gene during 5-aminolevulinic acid formation
in Pseudomonas aeruginosa."
RL J. Bacteriol. 177:1435-1443(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.L., Goltzy L., Tolentino E., Wesbrock-Wadman S., Yuan Y.,
RA Brody L.A., Goltzer S.N., Felger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Saiter M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

AC P21296;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flba protein.
GN FLBA OR CCI457.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=984422451; PubMed=9748431;
RA Lecerc G., Wang S.P., Ely B.;
RT "A new class of Caulobacter crescentus flagellar genes."
RL J. Bacteriol. 180:5010-5019(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Lamb M.T., Paulsen I.T., Nelson K.E.,
RA Eissen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Debey R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uteberck T., Tian K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN [3]
RP SEQUENCE OF 1-56 FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=91033011; PubMed=1699845;
RA Schenlein P.V., Gallman L.S., Winkler M.E., Ely B.;
RT "Nucleotide sequence of the Caulobacter crescentus flaf and flbr
genes and an analysis of codon usage in organisms with G + C-rich
genomes."
RL Gene 93:17-25(1990).
CC -----
CC -1- SIMILARITY: CONTAINS 4 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

RESULT 39
FLBA CAUCR
ID FLBA_CAUCR STANDARD; PRT; 596 AA.

QY 74 LAKVQASEDATHTQKARLFKALIELNPKDMQSYMDYGFVLVQKGLSGALIVFDKPSR 133
 DB 139 LGTFLCNIGDAGSIV---FFDESRLAPDFSKAVHNRFAFLRDJGEIETALADCEAMR 255
 QY 134 AIGYGRVVAIENMAYIY-----HGYAAKSPTRKD 165
 DB 256 SPG-----SPDLAMQFARATILLAGRVGEVAYESRPSPLSD 297

RESULT 40

COLA_VIBAL STANDARD; PRT; 814 AA.
 ID COLA_VIBAL

AC P43154;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Microbial collagenase precursor (EC 3.4.24.3).
 OS Vibrio alginolyticus.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxId=663;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RC STRAIN=Chemoovar. Iophagus;
 RX MEDLINE=92162015; PubMed=1311172;
 RA Takeuchi H., Shibano Y., Morihara K., Fukushima J., Inami S.,
 RA Kell B., Gilles A.-M., Kawamoto S., Okuda K.;
 RT "Structural gene and complete amino acid sequence of Vibrio
 alginolyticus collagenase.";
 RL Biochem. J. 261:703-708(1992).

-1- CATALYTIC ACTIVITY: Digestion of native collagen in the triple
 helical region at Xaa-[Gly bonds, with synthetic peptides, a
 preference is shown for Gly at P3 and P1'; Pro and Ala at P2 and
 P2'; and hydroxyproline, Ala or Arg at P3'.

-1- SUBCELLULAR LOCATION: Secreted.
 -1- PTM: THREE DIFFERENT ACTIVE FORMS OF THE COLLAGENASE ARE ASSUMED
 TO BE CLEAVAGE PRODUCTS DERIVED FROM THE SAME SIMPLE POLYPEPTIDE
 CHAIN.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M9.

-1- SIMILARITY: CONTAINS 1 PKD DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

CC EMBL; X62635; CAA44501.1; -.
 DR MEROPS; M09.001; -.
 DR InterPro; IPR002169; Miccollptase.
 DR InterPro; IPR000601; PKD_domain.
 DR InterPro; IPR000130; Zn_Mtpeptidase.
 DR Pfam; PF00801; PKD; 1.
 DR Pfam; PF01752; Peptidase M9; 1.
 DR PRINTS; PR00931; MICCOLLPTASE.
 DR SMART; SM00089; PKD; 1.
 DR PROSITE; PS50093; PKD; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydroxylase; Metalloprotease; Zinc; Zymogen; signal.
 FT SIGNAL 1 21
 FT PROPEP 22 75
 FT CHAIN 76 814
 FT DOMAIN 609 697
 FT METAL 477 477
 FT ACT_SITE 478 478
 FT METAL 481 481
 FT ZINC 481 481
 SQ SEQUENCE 814 AA; 89962 MW; E0E19C084D2185E CRC64;

Query Match 7.9%; Score 82; DB 1; Length 814;
 Best Local Similarity 26.2%; Pred. No. 22;
 Matches 44; Conservative 29; Mismatches 57; Indels 38; Gaps 11;

QY 30 NPOLAQI---RTQIAISLDMGKLDQAKQOUAALSADRF--APAVTLAKVQASEDA 84
 DB 253 NEQFVQIIGNQTDIAKALGDPA-----LAASSIGADEFMANAGRELGRLTKYTGNA 305
 QY 85 THQTKAQ--RLFEK-----AIELNPKDMQSY---MDYGFVLVQKGLSGALI---Y 127
 DB 306 SSVVKSQLSRIFEQYEMVGRGDVAVLAAADPASYYADCSBFGICNFET-ELKGLVLSQTY 364
 QY 128 FDKP-----SRAIGYGRVVAIENMAYI--YYHQ-YEAAKSPTRDDVN 167
 DB 365 TCSPTIRILSONMTQEOHAAACSKMGYHEGYFHOSLETGEQPVQDDHN 412

Search completed: July 8, 2003, 11:31:52
 Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 8, 2003, 11:27:39 ; Search time 78 Seconds
(without alignments)
538.892 Million cell updates/sec

Title: US-10-069-544-2

Perfect score: 1043
Sequence: 1 MKIRVKPMVAMGLVLSAC.....EIKSYDKLSDYKLPDYK 204

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protein:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151.5	14.5	224	2	O9ZH15
2	136.5	13.1	249	16	O8ZCT4
3	136.5	13.1	311	16	O8XIX5
4	136.5	12.7	379	17	O26186
5	124.5	11.9	1151	5	O18158
6	123	11.8	181	16	O9CJ37
7	122	11.7	342	16	O67735
8	120.5	11.6	329	16	O8XKU9
9	117.5	11.3	253	16	O9JZ41
10	117.5	11.3	253	16	O9JZ41
11	117.5	11.3	1046	4	O8WMM8
12	117.5	11.3	1046	11	O91Y38
13	116	11.1	1011	5	O91Y38
14	116	11.1	1059	5	O9V3X6
15	115.5	11.1	305	16	O97DM4
16	115	11.0	237	16	O9KRM4

17	114.5	11.0	1885	17	O8TQD1	O8TQD1 methanosaarc
18	114	10.9	931	10	O8RVB2	O8RVB2 lycopersico
19	113.5	10.9	513	17	O979D5	O979D5 thermoplasm
20	113	10.8	162	16	O67941	O67941 aquifex aeo
21	112.5	10.8	529	17	O9H455	O9H455 thermoplasm
22	112	10.7	932	10	O82039	O82039 petunia hyb
23	111.5	10.7	403	17	O26176	O26176 methanobact
24	111	10.6	252	16	O9HXJ2	O9HXJ2 pseudomonas
25	111	10.6	538	16	O9X294	O9X294 thermotoga
26	111	10.6	609	16	O92KK7	O92KK7 rhizobium m
27	109.5	10.5	311	16	O8TQD1	O8TQD1 anabaena sp
28	109	10.5	295	16	O92M02	O92M02 rhizobium m
29	109	10.5	658	16	O981X1	O981X1 rhizobium 1
30	108.5	10.4	614	3	O13499	O13499 podospira a
31	108	10.4	551	16	O981X2	O981X2 rhizobium 1
32	107.5	10.3	255	16	O8YD43	O8YD43 brucella me
33	107.5	10.3	400	17	O8TR22	O8TR22 methanosaarc
34	107	10.3	914	17	O8TUB5	O8TUB5 methanosaarc
35	106.5	10.2	212	16	O92BK3	O92BK3 listeria in
36	105	10.1	198	2	O51526	O51526 pseudomonas
37	104.5	10.0	707	16	O8YQD7	O8YQD7 anabaena sp
38	104	10.0	628	16	O8R147	O8R147 fuscobacteri
39	103.5	9.9	466	17	O8TNN0	O8TNN0 methanosaarc
40	103.5	9.9	944	10	O82422	O82422 hordeum vul
41	103	9.9	240	17	O97CQ8	O97CQ8 thermoplasm
42	103	9.9	280	16	O8YWD6	O8YWD6 anabaena sp
43	103	9.9	659	16	O8UHP7	O8UHP7 agrobacteri
44	103	9.9	744	5	O16376	O16376 caenorhabdi
45	102.5	9.8	236	16	O8YNN8	O8YNN8 anabaena sp

ALIGNMENTS

RESULT 1

ID O9ZH15 PRELIMINARY; PRT; 224 AA.
AC O9ZH15;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Hypothetical 25.2 KDa protein.
OS Desulfotomaculum thermocisternum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Peptococcaceae; Desulfotomaculum.
OX NCBI_TaxID=42471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST90;
RA Larsen O., Lien T., Birkeland N.-K.;
RT "Disimilatory sulfite reductase from Archaeoglobus profundus and
RT Desulfotomaculum thermocisternum: phylogenetic and structural
RT implications from gene sequences.";
RL Extremophiles 0:0-0(1999).
DR EMBL; AF074336; AAC96110.1; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 5.
DR SMART; SM00028; TPR; 5.
KW Hypothetical protein.
SQ SEQUENCE 224 AA; 25210 MW; 36021D7A609195A3 CRC64;

Query Match 14.5%; Score 151.5; DB 2; Length 224;
Best local similarity 23.3%; Pred. No. 3.5e-05;
Matches 48; Conservative 36; Mismatches 69; Indels 53; Gaps 5;

OY	28	KNNPQLAQRITQIAISLIDMGKLDQKQDALSADRFAPAYRTAKVQASDAQD	87
DB	19	QSNFGCATAKYNLGVLMLEQGRLEBAKTLFBEALADGTRMEAYVNLGIFYKCNLEKV	78
OY	86	TKAQR-----FKRIENPKDMQSYNDYGFIV	116
DB	79	ECNRLLAIVQIEPRYARGYANIGFYVLOWEKPEDEALIEVLOKAIENPRLIOQMCNLNAYVL	138

```

QY 117 OMGDLGSL-----IYDKSRALGYGRVVAIENMAYIYHOYEAKSPKTDYNNAKS 171
DB 139 QKQDLDRALSTNTLKMELIAPDFSLGH-----NNLAIYVY-----LKQDNRAAE 182
QY 172 ALERALLISGTO-HDEIKSYDKLLSD 196
DB 183 HKQALELGFENHPEFLKEMEPYLN 208

RESULT 2
Q8BCT4 PRELIMINARY; PRT; 249 AA.
AC 08BCT4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative fibrinial biogenesis protein.
GN YPO2881.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia
OC NCBI_TaxID=632;
RN 11]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titchell R.W., Holden M.T.G.,
RA Prentice M.B., Sebatina M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford A.V.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,
RT "Genome sequence of Yersinia pestis, the causative agent of plague.",
RT Nature 413:523-527(2001).
RL EMBL; AJ14154; CAC92132.1; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 3.
DR SMART; SM00028; TPR; 3.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 249 AA; 27499 MW; A14AD035A8685FA7 CRC64;

Query Match 13.1%; Score 136.5; DB 16; Length 249;
Best Local Similarity 26.7%; Pred. No. 0.00077;
Matches 48; Conservative 30; Mismatches 83; Indels 19; Gaps 7;

QY 1 MKIRVWPMVMAMGLVLSACQSTPIPKNNPQ--LAQIRTOAISLLMGKLDQAKQID 58
DB 1 MKITKLMRVCLVVS-VLTGCGST--PPENTSCAVAGQTRLOGLAVLAQGDILPAARKNLE 57
QY 59 AALSADROFAPAVRT-LAKYQASEDATHTQTKAQRLEFKAIEINPKDMQSYMDYGFYVQ 117
DB 58 KAVEADPQ--DYRTQGMATFYKORIGENSABQR-YQQMKLAPNGVLYNNYGAFLCS 113
QY 118 MGDLGALYFDPKPSRAIGYGRVVAIENMAYIYHOYEAKSPKTDYNNAKSALERAL 177
DB 114 LGQYVAQQQFSAALLPQYGVADSLNAGYFLR-----ANQKQKARVLSRL 164

RESULT 3
Q8XIX5 PRELIMINARY; PRT; 311 AA.
AC 08XIX5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein CPE1988.
GN CPE1988.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OC NCBI_TaxID=1502;
RN 11]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.",
RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RL EMBL; AF003192; BAB81694.1; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 8.
DR SMART; SM00028; TPR; 6.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 311 AA; 36424 MW; 67474FE0C40BA9B CRC64;

Query Match 13.1%; Score 136.5; DB 16; Length 311;
Best Local Similarity 25.9%; Pred. No. 0.001;
Matches 51; Conservative 29; Mismatches 64; Indels 53; Gaps 5;

QY 40 IAIISLDMGKLDQAKQIDALASADROFAPAVRTLAKYQASEDATHTQTKAQRLEFKAIE 99
DB 50 IALMLDELGTETKAKKYEALISNFEVRAVYGLAVYDERD---YKAIELYKKAID 106
QY 100 INPKDMQSYMDYGFYVQMGDLGALYFDPK----- 131
DB 107 INPVDRALIFLANLMDVNGKEGALIEYKTIYCKEDPFAHINTLSIYERDMLQRAY 166
QY 132 --SRAIGYEG-RVAIENMAYIYHOYEAKSPKTDYNNAKSALERALLISGTOHDEIK 187
DB 167 RWFSEKALQIDGHHYALFLNMGVY-----KRLNVDYDKAKKYEKA-----IK 208
QY 168 KSYDKLSLDYKLLSDYK 204
DB 209 KNFGYAFSFLNLAVTYK 225

RESULT 4
Q26186 PRELIMINARY; PRT; 379 AA.
AC 026186;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE O-linked GlcNAc transferase.
GN MTH83.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OC NCBI_TaxID=187420;
RN 11]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=93771463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldegrete T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Carno A., Bush D., Safer H., Petwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.W.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.",
RT J. Bacteriol. 179:7135-7155(1997).
RL EMBL; AF000799; AAB84589.1; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 10.
DR SMART; SM00028; TPR; 10.
KW Transferase; Complete proteome.
SQ SEQUENCE 379 AA; 43975 MW; 476138AB9E83E7 CRC64;

```

```

Query Match 12.7%; Score 132.5; DB 17; Length 379;
Best Local Similarity 26.9%; Pred. No. 0.0029;
Matches 49; Conservative 36; Mismatches 70; Indels 27; Gaps 7;

```

```

QY 28 KNNPOLAQRTOIAISLDMGKLDQAKQOOLDAALSADROFAPAVRTLAKVQASDATHQ 87
DB 78 KNNPFLAAMNNKGLVLKELGRYDEALBCEYKALINPKYAGAMNNKALVY---KELARY 134
QY 88 TKAQRLFEKALIELNPKDMOSYMDYGFYLVOMGDLGALIFPKP-----SRAIGYGR 140
DB 135 DEALECEYKALQINPKLADAWYKGSVLILYKRYKALKCEFEKALIELNPKYRFGWTGK- 193
QY 141 VVAIENMAYIYHYOYEAK-----SPTKDYNNAKSALERALISGTQGHDEIKKSYDK 192
DB 194 -ITLNNK-IT---BEALKCYDKVLQINPDDKAMNNKGLVNNEL---GRYDESLEYEK 245
QY 193 LL 194
DB 246 AL 247

RESULT 5
O18158 PRELIMINARY; PRT; 1151 AA.
AC O18158; Q21232;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE UDP-N-Acetylglucosamine--peptide N-acetylglucosaminyltransferase
DE (EC 2.4.1.1-) (O-GLCNAC) (OGT).
GN K04G7.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodietinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 321-340 AND 1060-1076, AND SUBCELLULAR
RP LOCATION.
RX MEDLINE=97238870; PubMed=9083068;
RX Lubas W.A., Frank D.W., Krause M., Hanover J.A.;
RT "O-linked GlcNAc transferase is a conserved nucleocytoplasmic protein
RT containing tetratricopeptide repeats.";
RL J. Biol. Chem. 272:9316-9324 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RX Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RX Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
RX Craxton M., Dear S., Du Z., Durbin R., Faveille A., Fraser A.,
RX Fulton L., Gardner A., Green P., Hawkins T., Haller L., Jier M.,
RX Johnston L., Jones M., Kershaw J., Kilsten J., Laister N.,
RX Lacroille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RX Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RX Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RX Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RX Waterston R., Watson A., Weinstock L., Wilkinson-Spratol J.,
RA Woldman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RT Nature 368:32-38 (1994).
RN [3]
RP REVISIONS.
RP Waterston R.;
RL Submitted (May-2002) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: ADDITION OF NUCLEOTIDE-ACTIVATED SUGARS DIRECTLY ONTO
CC THE POLYPEPTIDE THROUGH O-GLYCOSIDIC LINKAGE WITH THE HYDROXYL OF
CC SERINE OR THREONINE.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + PEPTIDE = UDP +
CC N-ACETYL-BETA-D-GLUCOSAMINYL-PEPTIDE.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND PERINUCLEAR.
CC -1- SIMILARITY: CONTAINS 12 TPR REPEATS.
DR EMBL, U77412; AAB63465.1; -
DR EMBL, U21320; AAB62535.2; -
DR WormPep; K04G7.3; CE25042.

```

```

DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 12.
DR SMART; SM00028; TPR; 11.
KW Transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat.
FT REPEAT 193 226 TPR 1.
FT REPEAT 227 260 TPR 2.
FT REPEAT 261 294 TPR 3.
FT REPEAT 295 328 TPR 4.
FT REPEAT 329 362 TPR 5.
FT REPEAT 363 396 TPR 6.
FT REPEAT 397 430 TPR 7.
FT REPEAT 431 464 TPR 8.
FT REPEAT 465 498 TPR 9.
FT REPEAT 499 532 TPR 10.
FT REPEAT 533 566 TPR 11.
FT REPEAT 567 577 TPR 12 (INCOMPLETE).
FT DOMAIN 63 106 GLN-RICH.
FT DOMAIN 591 607 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 1151 AA; 127981 MW; D1C8F669401C55B CRC64;

Query Match 11.9%; Score 124.5; DB 5; Length 1151;
Best Local Similarity 27.3%; Pred. No. 0.061;
Matches 44; Conservative 28; Mismatches 64; Indels 25; Gaps 6;

QY 30 NPOLAQIRTOIAISLDMGKLDQAKQOOLDAALSADROFAPAVRTLAKVQASD---ATH 86
DB 257 NPDLYCRSDGNLKKMGRLSEAKVCYKALFTQPOFAVAMSNLGCYFNSQGHIMLAIH 316
QY 87 QTKAQRLEKALIELNPKDMOSYMDYGFYLVOMGDLGALIFDKPSRAIGYGRVAIE- 145
DB 317 H-----FEKAVTLDPNFLDYLIVIMIGVNLKARIFDRAVSAY---LRALINSGHNAVHG 367
QY 146 NMAYIYHYO-----YEAA---KSPTKDYNNAKSALE 174
DB 368 NLACVYIEQGLIDALIDTYKALIDLPHPFPAVCNLANALK 408

RESULT 6
O9CJU7 PRELIMINARY; PRT; 181 AA.
AC O9CJU7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical protein PM2008.
GN PM2008.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RX May B.J., Zhang Q., Li L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AE006237; AAK04092.1; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 3.
DR SMART; SM00028; TPR; 3.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 181 AA; 20911 MW; E756709F73342A2D CRC64;

Query Match 11.8%; Score 123; DB 16; Length 181;
Best Local Similarity 26.6%; Pred. No. 0.0072;
Matches 45; Conservative 25; Mismatches 77; Indels 22; Gaps 5;

QY 15 LVISACSTPIPPKNP-----QLAQIRTOIAISLDMGKLDQAKQOOLDAALSADROFAP 69
DB 15 LFLSAC---VSQSGNTPFSHEAKARIELGLAYLSQHPFARAKQNFDKALSHAPNYLL 70
QY 70 AVRTLAKVQASDATHQTKAQRLEKALIELNPKDMOSYMDYGFYLVOMGDLGALIFYD 129

```

```

Db      71 SHAVALTYLYOKO---CHLDLARAHAKEKALIKLNDKQSDVANNIGAFICQCGSEFQAYKQPS 127
QY      130 KPSRAIGYEGRAVAIENMAY-----IYHOYEAKSPTKDYNNAK 170
Db      128 QALNAPHYYQHADTYENLWVCASFADQYRYOQ--NLQHLTQDPDPRAK 174

RESULT 7
ID 067735 PRELIMINARY; PRT; 342 AA.
AC 067735;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical protein AQ_1896.
GN AQ_1896.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_Taxid=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Anujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
DR EMBL: AE000762; AAC07708.1; -.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 5.
DR SMART: SM00028; TPR; 5.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 342 AA; 40375 MW; 99DDF39CF2956DE9 CRC64;

Query Match 11.7%; Score 122; DB 16; Length 342;
Best Local Similarity 24.0%; Pred. No. 0.02;
Matches 43; Conservative 36; Mismatches 64; Indels 36; Gaps 8;

QY 30 NPQLAQITQIAISLDMGKLDQAKQDLAALSAD--RQFAPAYRTLAKYQASDATHQ 87
Db 89 NNPYSARKNLGLTYKLGRIYERALKYLGQANDREYKEKHEAFYLLAKYVEKQDLKRY 148
QY 88 TKAQRLFEKALIELNPKMQSYNDYGYFLVQMGDLSCA-LIY-----FDKSRALGYEG 139
Db 149 V---RYLEKAVAVNPNFVQAQLELAQAYENLGYEAEKIKYKSLNLNGFNKP--FLKYK- 202
QY 140 RVAIEMNAVYVHOYEAKSPKRDYNNAKSLERAL-----ISGTQHEIKKSYNKL 194
Db 203 -----LAEVY-----KSGDYERAREIKELLYKENTNTQREKVELLTKVL 245

RESULT 8
Q8XKU9 PRELIMINARY; PRT; 329 AA.
AC Q8XKU9;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein CPE1293.
GN CPE1293.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_Taxid=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Onohima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;

```

```

RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater."
RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
RL EMBL: AP003190; BAB80999.1; -.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 7.
DR SMART: SM00028; TPR; 7.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 329 AA; 38675 MW; FB4CB8A104FD72158 CRC64;

Query Match 11.6%; Score 120.5; DB 16; Length 329;
Best Local Similarity 24.4%; Pred. No. 0.026;
Matches 43; Conservative 40; Mismatches 74; Indels 19; Gaps 6;

QY 30 NPQLAQITQIAISLDMGKLDQAKQDLAALSADROAPAYRTLAKYQASDATHQ 89
Db 160 NPYFNALISGRAYCAFEFGDYLATKYLYQYEVSIKNNAFYKLGCECYFLINDKNSK 219
QY 90 AQRLEKALIELNPKMQSYNDYGYFLVQMGDLSCALYEDKPSRAIGYEGRAVAIENMAY 149
Db 220 ---MYEKLEIDQENDASYCYANVLCG-----HYDK---ALNNLNKALISPNYF 266
QY 150 IYHOYEAKSPTKDYNNAKSALERALISGTQHEIKK--SYDKLSDYKLLSDY 203
Db 267 MAFY-YKARAYSLKKDPFNAYFYLERAF---NYCDLSLEIALKQEMLTNLKFFSKF 318

RESULT 9
Q9JZ41 PRELIMINARY; PRT; 253 AA.
AC Q9JZ41;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Fimrial biogenesis and twitching motility protein, putative.
GN NMB3309.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=2015755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Risen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn S.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Uitterback T.R., Khouli H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
DR EMBL: AE002479; AAF41684.1; -.
DR TIGR: NMB3309; -.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 3.
DR SMART: SM00028; TPR; 3.
KW Complete proteome.
SQ SEQUENCE 253 AA; 28424 MW; 6BCB1314D61A74DA CRC64;

Query Match 11.3%; Score 117.5; DB 16; Length 253;
Best Local Similarity 23.6%; Pred. No. 0.033;
Matches 43; Conservative 42; Mismatches 78; Indels 19; Gaps 5;

QY 4 RYKWPWAMAGVLVSAQST--PIPPKNNPQLAQITQIAISLDMGKLDQAKQDLAAL 61
Db 8 RISTLLVATLG---ACSTYRPSRAEKANYSNIKTQLAMWYMGQYRQATASIEDAL 63
QY 62 SADROFAPAYRTLAKYQASDATHQYKAQRLFEKALIELNPKMQSYNDYGYFLV-QMGD 120
Db 64 KSDPKNELAMLVRAIEITYLKK--VNDKAGSSPFGALSIKSDSAEINNINNYGMLCGRLNR 120

```



```

QY 127 YFDKSRRAIGYGRVVAIF-----NNAYIYHQ-----YEA---KSPYKDY 166
DB 241 F----DRAVAAYLRALSLSPNNAVHGNLACVYEGGLDLDIDYRRAIELQHPFPAY 296
QY 167 NNAKSALE 174
DB 297 CNLANALK 304

RESULT 13
QY148 PRELIMINARY; PRT; 1011 AA.
AC QY148;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE BCDNA:GH04245 (Fragment).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA Agbayani A., Arcina T.T., Baxter E., Blazey R.G., Butenhoif C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Fafian D.E., Frise E.,
RA Galle R., George R.A., Harris N.W., Hoskins R.A., Evans-Hoim M.,
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Paclob J.M.,
RA Park S., Segneita A., Sethi H., Snit E., Svirkas R.R., Weinburg T.,
RA Ceiniker S.E.;
RT "Full length Drosophila melanogaster cDNA sequence."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF145622; AAD38597.1; -
DR FlyBase; FBgn0040295; Ogt.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 11.
DR SMART; SM00028; TPR; 11.
FT NON TER 1011 1011
SQ SEQUENCE 1011 AA; 113021 MW; 0156474FF11EF5A3 CRC64;

Query Match 11.1%; Score 116; DB 5; Length 1011;
Best Local Similarity 24.8%; Pred. No. 0.27; Mismatches 32; Gaps 7;
Matches 50; Conservative 37; Indels 83;

QY 18 SACSSTPIPKNNPOLAQRTOAISLDMGKLDQAKQOLDALASDRQAPAPAYRTLAKV 77
DB 170 SAVQAYITLQYNPDLYCVASDIGNLTKALGRLEAKACVLRKATETCPGFAVAMSWLGCV 229
QY 78 YQASED---ATHQTKAQRLEFEKAIEINPKDMOSYMDYGFVLVQMGDLGALLYFDPKPSRA 134
DB 230 FNAQGEIWLAIHH-----FEKAVTLDPNPLDLAINLGNVLKEARIFDAVAAY---LRA 280
QY 135 IGGE-GRVAIEINMATIYHQ-----YEA---KSPYKDYNNAKSALERALLISGT 181
DB 281 LNISPNNAVHGNLACVYEGGLDLDIDYRRAIELQHPFPAYCNLANALK----- 333
QY 182 QHDEIKSYDKLSDYKLSYD 203
DB 334 EKGQVKEAECCVNTALRLCSNH 355

RESULT 14
QY3X6 PRELIMINARY; PRT; 1059 AA.
AC QY3X6;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE BCDNA:GH04245 protein.
OGT OR BCDNA:GH04245 OR CG10392.

QY 18 SACSSTPIPKNNPOLAQRTOAISLDMGKLDQAKQOLDALASDRQAPAPAYRTLAKV 77
DB 170 SAVQAYITLQYNPDLYCVASDIGNLTKALGRLEAKACVLRKATETCPGFAVAMSWLGCV 229
QY 78 YQASED---ATHQTKAQRLEFEKAIEINPKDMOSYMDYGFVLVQMGDLGALLYFDPKPSRA 134
DB 230 FNAQGEIWLAIHH-----FEKAVTLDPNPLDLAINLGNVLKEARIFDAVAAY---LRA 280

```

```

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiza S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M.E., Kadish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matcel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefel F., Shen T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong R.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Rubenstein D.S., McEwen D.G., Peifer M.A.;
RT "Drosophila melanogaster O-glycosyltransferase."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003787; AAF57318.1; -
DR EMBL; AF217788; AAF52311.1; -
DR FlyBase; FBgn0040295; Ogt.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 11.
DR SMART; SM00028; TPR; 11.
FT Transferase.
SQ SEQUENCE 1059 AA; 118649 MW; 48311D2370321B7D CRC64;

Query Match 11.1%; Score 116; DB 5; Length 1059;
Best Local Similarity 24.8%; Pred. No. 0.29; Mismatches 32; Gaps 7;
Matches 50; Conservative 37; Indels 83;

QY 18 SACSSTPIPKNNPOLAQRTOAISLDMGKLDQAKQOLDALASDRQAPAPAYRTLAKV 77
DB 170 SAVQAYITLQYNPDLYCVASDIGNLTKALGRLEAKACVLRKATETCPGFAVAMSWLGCV 229
QY 78 YQASED---ATHQTKAQRLEFEKAIEINPKDMOSYMDYGFVLVQMGDLGALLYFDPKPSRA 134
DB 230 FNAQGEIWLAIHH-----FEKAVTLDPNPLDLAINLGNVLKEARIFDAVAAY---LRA 280

```

```

QY 135 ICYE-GRVVAIENMAYIYHQ-----YEA---KSPTPDYNNASALERALISGT 181
DB 281 LNLSPNNNAVHGNLACVVEEGGLDLATITRYRAIELOPNPDVCANLANK----- 333
QY 182 QHDEIKSYDKLSIDYKLSIDY 203
DB 334 EKGQVKEADCVYNTALRLCSNH 355

RESULT 15
Q97DM4 PRELIMINARY; PRT; 305 AA.
AC 097DM4;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE TPR-repeat-containing protein.
GN CAC3449.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_Taxid=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=2159325; PubMed=1146286;
RA Neelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabete F., Doucet-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.,
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.",
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007843; AAK81379.1; -
DR InterPro: IPR003975; Shal_channel.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 5.
DR PRINTS: PR01497; SHALCHANNEL.
DR SMART: SMO0028; TPR; 7.
KW Complete proteome.
SQ SEQUENCE 305 AA; 35938 MW; 8129C4E947C9EABC CRC64;

Query Match 11.1%; Score 115.5; DB 16; Length 305;
Best Local Similarity 25.5%; Pred. No. 0.063;
Matches 47; Conservative 35; Mismatches 61; Indels 41; Gaps 7;

QY 36 IRTQIAISL--DMGKLDQAKQDLALSADROFAPAYRTAKVYQASEDATHTQKQRL 93
DB 42 VDTIINLALYDLSLGEVKEAEYKEALSIDYERAYGLATYDEKED--YEAIVL 98
QY 94 PEKALELNDPKQMSYMDYGVYLVQMGDLSGALITFDKPSRAIGYGRVVAIENMAY---- 149
DB 99 YNKAIYINNHYKAV---FFLANAYDLSG-----QKXLALEYEKLSLDDMFWSNL 148
QY 150 ----IYVHQ-----YEAASPTKDDYNN-----AKSALERLISGTQHDEIK 187
DB 149 NIGCIYEENKNDLAVRLFSKALKINPNHNLAPMGVYCCFNMIEKAINYEKSIENK 208
QY 188 KSYD 191
DB 209 KSYE 212

RESULT 16
Q9KRM4 PRELIMINARY; PRT; 237 AA.
AC 09KRM4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Fimbrial biogenesis and twitching motility protein, putative.
GN VC1612.
OS Vibrio cholerae.

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Hedeberg J.F., Eiken J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.",
RL Nature 406:477-483(2000).
DR EMBL: AE004238; AAF94766.1; -
DR TIGR; VC1612; -
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 4.
DR SMART: SMO0028; TPR; 4.
KW Complete proteome.
SQ SEQUENCE 237 AA; 26649 MW; DECFA17GAC067F4 CRC64;

Query Match 11.0%; Score 115; DB 16; Length 237;
Best Local Similarity 23.5%; Pred. No. 0.05;
Matches 50; Conservative 40; Mismatches 63; Indels 60; Gaps 11;

QY 10 VMANGVL-----SACQSTPIPKNNPQALQIRTOAISLMDGKLDQAKQ 56
DB 4 VFGIGLITIALAGCVTETAGNATQSDP-----TEMAEARIALGLGYLENGSMIKAREN 57
QY 57 LDALASADROFAPAY-----RLTAYQASEDATHTQKQRFKRAIELNDPKQMSYMDY 112
DB 58 LEKAL---QRAPIYRSQSLMAHYEA---VGENDSRKRYRRLSEHPKGNVLANFG 110
QY 113 FYLVQMGDLSGALITFDKPSRAIGYGRVVAIEN-MAYIYHQY-----AAKSPTKDY 166
DB 111 TFLCKQSHYDTRADQYFR-----AVEQPYLYLSASTENAGLCALYKAGKTD-- 156
QY 167 NNKASALERLISGTQHDEIKSYDKLSIDYKL 199
DB 157 -NAREYFKRAL-----DHP-----NRLLSITQL 179

RESULT 17
Q8TOD1 PRELIMINARY; PRT; 1885 AA.
AC 08TOD1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE TPR-domain containing protein.
GN MA1613.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_Taxid=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Deaerliano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tittell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Gues A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;

```

"The genome of Methanosarcina acetivorans reveals extensive metabolic RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE010832; AAM05026.1; -.
 KW Complete proteome.
 SQ SEQUENCE 1885 AA; 214339 MW; 6D152837737777AA2 CRC64;

Query Match 11.0%; Score 114.5; DB 17; Length 1885;
 Best Local Similarity 21.4%; Pred. No. 0.84;
 Matches 51; Conservative 36; Mismatches 74; Indels 77; Gaps 7;

QY 31 POLAQRTOIAISLDMGKLDQAKQALDPAALSADROFAPAYRTLAKYQASSEDATHQTKA 90
 DB 1591 PDYACWMLGASIELDYERALQALDLVLEREPAHDAIYNALVFNLEF---YEEA 1647
 QY 91 QRLFKALINPKMQSMQYMDYGFYLVQMGDISGALIYDKPS----- 132
 DB 1648 ARTEFQLEASPEDESINYLGLCILEDEDEKALKAEKALFNPKNEALYNAATTLI 1707
 QY 133 -----RAIGYGRVVAI-----ENMAYIYHOY-BAKS---PTKDDYNNAKS- 171
 DB 1708 KLNRIQESLGFDRLEISPENYDAMTKGVAFQMLEQYREALKSPNVLKQDPNNIKAV 1767
 QY 172 -----ALERRALISGTQH-----DEIKSYDKLL 194
 DB 1768 YNVGVCFQKLYETARAFAKELTLINPWEHPSLRVYGLSLAKTGVDYEDALKAKEKLL 1825

RESULT 18
 Q8RVB2 PRELIMINARY; PRT; 931 AA.
 AC Q8RVB2;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE SPY protein.
 GN SPY.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Assteridae; euasterids I; Solanales; Solanaceae; Solanum.
 CX NCBI_TaxId=4081;
 GN [1]
 RP SEQUENCE FROM N.A.
 RA Greb T., Schmitz G., Theres K.;
 RT "Spinally and Procera encode different components of the GA signal transduction pathway in tomato."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ312093; CA085168.1; -.
 DR EMBL; AJ312094; CA085169.1; -.
 SQ SEQUENCE 931 AA; 103491 MW; 1933A81C48132C62 CRC64;

Query Match 10.9%; Score 114; DB 10; Length 931;
 Best Local Similarity 22.5%; Pred. No. 0.37;
 Matches 47; Conservative 32; Mismatches 94; Indels 36; Gaps 6;

QY 14 GLVLSACQSTPIPPKNNPQ-----LAQRTQIAISLDMGKLDQAKQALDPAALSADRO 66
 DB 130 GRIVEAASEYKALKADPSYTPAABCLATLDTIGSLKAGTGGIGIKYEAATIDSH 189
 QY 67 FAPAYRTLAKYQASSEDATHQTKAQLFEKALINPKMQSMQYMDYGFYLVQMGDISGALI 126
 DB 190 YAPAYTNLGVVY---SEMQYDMLNCEYKALERPMVAEACNMGVIFPKNGDLESALA 246
 QY 127 YPDK-----PSRAIGYGRVVAIENMAYIYHOYEAKSPKDDYNNAKSALERALI--- 178
 DB 247 CYERCLAVSPNFEIANKNNALATLD-----GTAKLGGDINGVAYYKALKALCINW 297
 QY 179 -----SGTQHEIKSYDKLLSDYKL 199
 DB 298 HYADAMYNLGVAYGEMLK-FDMAIVFYEEL 325

RESULT 19
 Q979D4 PRELIMINARY; PRT; 513 AA.
 ID Q979D4
 AC Q979D4;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Hypothetical protein TV1227.
 GN TV1227 OR TVG1265169.
 OS Thermoplasma volcanium.
 OC Archaea; Euryarchaeota; Thermoplasmatata; Thermoplasmales;
 OC Thermoplasmataceae; Thermoplasma.
 CX NCBI_TaxId=50339;
 GN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GS1 / DSM 4299 / JCM 9571;
 RX MEDLINE=20570466; PubMed=11121031;
 RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
 RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Kawamoto T.,
 RA "Archaeal adaptation to higher temperatures revealed by genomic RT sequence of Thermoplasma volcanium."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
 DR EMBL; AP000995; BAB60369.1; -.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 6.
 DR SMART; SM00028; TPR; 7.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 513 AA; 58475 MW; 6865C491168B1701 CRC64;

Query Match 10.9%; Score 113.5; DB 17; Length 513;
 Best Local Similarity 24.7%; Pred. No. 0.18;
 Matches 44; Conservative 28; Mismatches 67; Indels 39; Gaps 6;

QY 28 KNNPOLAQRTOIAISLDMGKLDQAKQALDPAALSADROFAPAYRTLAKYQASSEDATHQ 87
 DB 76 KNNPELFSLK---SICYQSMQDPFRKSEADKAIKAKQSPFAYYNNAAALRGIGDV--- 129
 QY 88 TKAQRLFEKALINPKMQSMQYMDYGFYLVQMGDISGALIYFDRKSAIGYGRVVAIENM 147
 DB 130 DGAERDFFRYEIMOPSPDPHYDL-----AEIYFEKKD----- 162

QY 148 AYIYHOYEAKSPKDDYNNAKSALERALISGTQHEIKSYDK-LLSDYKLLSDYK 204
 DB 153 ---YKKAEEYKAVLRKDKSNERKAYDILQIHLATND--GGYFKALLEAFKOTENFK 215

RESULT 20
 Q67941 PRELIMINARY; PRT; 162 AA.
 ID Q67941
 AC Q67941;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Hypothetical protein A0_2197.
 GN A0_2197.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 CX NCBI_TaxId=63363;
 GN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.B., Overbeek K., Sneed M.A., Keller M., Anjlay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex RT aeolicus."
 RL Nature 392:353-358(1998).
 DR EMBL; AE000777; AAC07909.1; -.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 4.
 DR SMART; SM00028; TPR; 3.

KW Hypothetical protein, Complete proteome.
SQ SEQUENCE 162 AA; 19020 MW; 446C0BB9B9BFD517 CRC64;

Query Match 10.8%; Score 113; DB 16; Length 162;
Best Local Similarity 28.3%; Pred. No. 0.045;
Matches 39; Conservative 18; Mismatches 49; Indels 32; Gaps 4;

QY 47 MGKLDQAQQLDAAASADRPAPAYRTLLAKVYQASEDAT-----HOTKAQRLPEK 96
DB 24 LGDDLKALE-----LYRKSIDVPTAEAWTFLEGMAYSMRNYEGALBACKR 69
QY 97 AIELNPKDMOSYMDYGFYLVQMGDLSGALIFDKPSRAIGYGRVAIENMAYIYHQ-- 154
DB 70 AIEIDPDFGNPYNDIGSYLIEGKLEDEALEWLEKAKRARKYEPHPYPIINLAKVYMLQCK 129
QY 155 -YEAKS-----PTKDDY 166
DB 130 LYEALEIEIEAIKIKYD 147

RESULT 21

Q9HJ45 PRELIMINARY; PRT; 529 AA.

AC Q9HJ45; 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
GN Hypothetical protein Tail27.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmales;
OC Thermoplasmales; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed11029001;
RA Ruepp A., Granel W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
Mewes H.-W., Fritsman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermophilic scavenger Thermoplasma
acidophilum.";
RL Nature 407:508-513 (2000).
DR EMBL; AL445066; CAC12254.1; -.
DR HSSP; P53041; 1A17.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 7.
DR SMART; SM00028; TPR; 6.
KW Hypothetical protein, Complete proteome.
SQ SEQUENCE 529 AA; 60538 MW; 2E71E6307755F868 CRC64;

Query Match 10.8%; Score 112.5; DB 17; Length 529;
Best Local Similarity 21.4%; Pred. No. 0.23; 57; Indels 57; Gaps 5;
Matches 40; Conservative 33; Mismatches 57; Indels 57; Gaps 5;

QY 28 KNNPOLAQRTOIAISLDMGKLDQAQQLDAAASADRPAPAYRTLLAKVYQASEDATQ 87
DB 91 KNNPBLISLK---SLYGSLLGKFNESKIKATKAKDPYPPAYVYRABAEQYLE---Y 144
QY 88 TTAQRLFEFAIELNPKDMOSYMDYGFYLVQMGDLSGALIFDKPSRAIGYGRVAIENM 147
DB 145 DIAKDLQKYLEWQPNPDPAVND-----L 168
QY 146 AVIYHYQVAAKSPTMDVNNNAKALERALISGTOHD-----BIKSYDKLSYD 197
DB 165 ADMEHHEGYSKKA-----LQHYNTAIKKKESTDAHDKLNTILKAKDIENYKLELEAF 223
QY 198 KLSLDPK 204
DB 224 KQTEDEK 230

RESULT 22

Q82039 PRELIMINARY; PRT; 932 AA.

AC Q82039; 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
GN SPINDLY protein.
OS *Pectunia hybrida* (Pectunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asterales; eusterids I; Solanales; Solanaceae; Pectunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. VR. TISSUE=ANTHERS;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y17720; CAA76834.1; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 10.
DR SMART; SM00028; TPR; 9.
SQ SEQUENCE 932 AA; 103939 MW; 4F0896631302D95D CRC64;

Query Match 10.7%; Score 112; DB 10; Length 932;
Best Local Similarity 22.5%; Pred. No. 0.54;
Matches 47; Conservative 32; Mismatches 94; Indels 36; Gaps 6;

QY 14 GLVLSACGSTRIPPNRQ-----LAQITQALISLDMGKLDQAQQLDAAASADQ 66
DB 130 GLVLAASGYKALKADSYRAECLAIVLTIGTSLKLGNSQEGIQKYEAIRKIDSH 189
QY 67 PAPAYRTLLAKVYQASEDATHOTKAQRLPEKALIELNPKDMOSYMDYGFYLVQMGDLSGALI 126
DB 190 YAPAYNNGVY---SEMQYDMALNCYKKAIERPMAEYCNMGVYIKRGRDESALA 246
QY 127 YFDK-----PSRAIGYGRVAIENMAYIYHQVEAASPTKDYNNNAKALERALI--- 178
DB 247 CYERCLAVSPNFKLNMAIALTDL-----GTVKLEGDINGQVAYYKALYYNM 297
QY 179 -----SGTHDEIKSYDKLSYK 199
DB 298 HYADAMYNLGVAYGEMK-FDMAIVFYEL 325

RESULT 23

O26176 PRELIMINARY; PRT; 403 AA.

AC O26176; 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
GN O-linked GlcNAc transferase.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
Harrison D., Hoang L., Keagle P., Lumm W., Potier B., Qiu D.,
Spadator R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
Jiwani N., Carno A., Bush D., Safer H., Patwell D., Prabhakar S.,
McDougal S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155 (1997).
DR EMBL; AE000796; AAB84576.1; -.
DR InterPro; IPR001440; TPR.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaute V., Masny D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsberger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
 RA "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021".
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL, AL591785; CAC45417.1; -.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 8.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 609 AA; 67667 MW; D771F3E911E69542 CRC64;

Query Match 10.6%; Score 111; DB 16; Length 609;
 Best Local Similarity 22.9%; Pred. No. 0.38;
 Matches 47; Conservative 35; Mismatches 71; Indels 52; Gaps 8;
 QY 27 PKNQALQIRTOAISLIDMKLDAQOQLDALSD-----RQF-----APAYRT 73
 DB 348 PESSWRRLSELQGLSLAGIGKDEAKKHLKGLIDVDPKXIRNYLAAYGSVLSDAKNYKE 407
 QY 74 LAKVQ-----ASEDATHTQKORLFEKATLNPKNQMSYMD 110
 DB 408 MGEIYDRAVAALIGVPKRSQWTVFQRGIAIYRQKLEKAEKFNKAELELNDPOQVILNY 467
 QY 111 YGFYLVQNG-DLSGALIFYDKPSRAIGYGRVALENNAYIYHOYEAKSPTKDDYNN 169
 DB 468 LGYSVNDKNNINLEDELMIKRAVELKPPDDGYIV--DSIGMAVFRN-----NRFDEA 516
 QY 170 KSALERA--LISG--TQHDEIKSY 190
 DB 517 VTELERAELMAGDATINDHLDAY 541

RESULT 27
 Q8YP20 PRELIMINARY; PRT; 311 AA.
 AC Q8YP20;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein A114382.
 GN A114382.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.,
 RA "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL; AP003596; BAB76081.1; -.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 7.
 DR SMART; SM00028; TPR; 7.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 311 AA; 34369 MW; B8D71881D8A8833 CRC64;

Query Match 10.5%; Score 109.5; DB 16; Length 311;
 Best Local Similarity 23.0%; Pred. No. 0.21;
 Matches 54; Conservative 37; Mismatches 81; Indels 63; Gaps 9;

QY 13 MELVLSACQSTPIPP-----KNPQALQIRTOAISLIDMKLDAQOQLDALSA 63
 DB 85 IGLALR--QTGQLDPADAFVRATQSDNPFALAFANIGSLLEGNNLQANDYIQLRALEL 142
 QY 64 DRQFAPAYRTLAKYQASED-----ATHQ-----TKQR 92
 DB 143 EPRGFAYNYGLVRQOQNMEGALASFOKAVELSKNAPBPHYLYGISYLOQKLINEKN 202
 QY 93 LFEKATLNPKNQMSYMDYGFYLVQNGDLSGALIFYDK-----PSRAIGYGRVALENN 147
 DB 203 AFNQHAKINPYSAYNYLVGLVFNQGPQALTRFKSAEAMPYPAATYG-----AGL 257
 QY 148 AYIYVHOY-EAKSPTKDDYNNAKSALERALISGTQHDEIKSYDKLSDYKLIS 201
 DB 258 VFTQNGYGEAKY-----FNAKKN-----LYSTQGNQMAKNMEQLQOVQNNIN 302

RESULT 28
 Q92M02 PRELIMINARY; PRT; 295 AA.
 ID Q92M02;
 AC Q92M02;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein R02857.
 GN R02857 OR SMC03935.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaute V., Masny D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsberger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
 RA "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021".
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591785; CAC47436.1; -.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 6.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 295 AA; 32288 MW; 9C6C2555E1E7F709 CRC64;

Query Match 10.5%; Score 109; DB 16; Length 295;
 Best Local Similarity 27.2%; Pred. No. 0.22;
 Matches 43; Conservative 18; Mismatches 75; Indels 22; Gaps 5;
 QY 35 QIRTOAISLIDMKLDAQOQLDA-----ALSADROFAPAYRTLAKYQASEDATHQ 87
 DB 140 QINPNYVAVAYGRNLYRQANQDLDAFNDFKATLELTADPRAVHNGILYQA--RNQH 196
 QY 88 TKQARLFEKATLNPKNQMSYMDYGFYLVQNGDLSGALIFYDKSRRAIGYGRV-ATEN 146
 DB 197 AQAIEDTSKALSTSPSSPEPYNGRISYVAGDDDNA--FSDFNATINLNGKLAESWAN 253
 QY 147 MAYIYVHOYEAKSPTK-----DDYNNAKSALER 176
 DB 254 QALYERGDGAKAKAKSYSHALSLDPRYEPRAALAA 291

RESULT 29
 Q98IX1 PRELIMINARY; PRT; 658 AA.
 ID Q98IX1;
 AC Q98IX1;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein ml12209.

```

GN ML2209.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxId=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099; PubMed=11214968;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Ref. 7:331-338(2000).
DR EMBL; AP002999; BAB49395.1;
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 14.
DR SMART; SM00028; TPR; 14.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 658 AA; 73392 MW; 52304A37D5BF3CD1 CRC64;

Query Match 10.5%; Score 109; DB 16; Length 658;
Best Local Similarity 21.6%; Pred. No. 0.62;
Matches 42; Conservative 32; Mismatches 66; Indels 54; Gaps 6;

QY 48 GKLDQAQQLDALSADRFAPAYRTLKYQASEDATHQKQRLFEKAIELNPKMQS 107
DB 218 GNLERAIADYSQAISLDAHAEAYRWADAWKRGD-TDQALSD-YTAIRLADPGDAET 274
QY 108 YMDGYFLVQMGDLSGALLYFDK-----PSRAIYEGR----- 140
DB 275 FRNARARWERGRDYDRAIADVDQAIAPADAVAYNGRGMWSLKHTDRAIVDYKATA 334
QY 141 -----VVAIENMAYIYHOYEAAKSPKDDYNNKSALEBALISGTHODEI----- 186
DB 335 FDPNYVALAYDNLGLAWM-----DKGLDLRAISFDQAVIVDPRIAPAYNDRLARM 385
QY 187 -KSYDYLSDYKL 199
DB 386 DKNOYDLAIADYNN 399

RESULT 30
013499 PRELIMINARY; PRT; 614 AA.
AC 013499;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE TOM70 protein.
GN TOM70.
OS Podospira anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Lasiosphaeriaceae; Podospira.
OX NCBI_TaxId=5145;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S.
RX MEDLINE=98001562; PubMed=9343397;
RA Jame-Verny C., Contamine V., Boulay J., Zickler D., Picard M.;
RT "Mutations in genes encoding the mitochondrial outer membrane proteins
RT Tom70 and Mdm10 of Podospira anserina modify the spectrum of
RT mitochondrial DNA rearrangements associated with cellular death.";
RL Mol. Cell. Biol. 17:6359-6366(1997).
DR EMBL; Y14750; CAAT5047.1;
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 10.
DR SMART; SM00028; TPR; 9.
DR TIGRPFAM; TIGR00990; 3a0801s09; 1.
SQ SEQUENCE 614 AA; 68145 MW; B7A3FCDD2A5013D2 CRC64;

```

```

Query Match 10.4%; Score 108.5; DB 3; Length 614;
Best Local Similarity 23.4%; Pred. No. 0.62;
Matches 36; Conservative 37; Mismatches 64; Indels 17; Gaps 4;

QY 28 KNPOLAQIRTOIAISLDMGKLDQAQQLDALSADRFAPAYRTLKYQASEDATHQ 79
DB 461 KNPKLPDYVYNYGELLDDQTFSEAIKFPDAILEKETEMSNVLPINKSLA-LFQ 519
QY 80 ASEDATHQKQRLFEKAIELNPKMQSYMDGYFLVQMGDLSGALLYFDKPSRAIYEG 139
DB 520 WKQDF-----SEAKLTEKALITDPECDIATWAOULLQOGVTEALKYFERAAELARTEG 576
QY 140 RVVAIENMAYIYHOYEAAKSPKDDYNNKSALEBALISGTHODEI----- 173
DB 577 ELVSA-----ISYATARTQIQOVQKRYDPLASKL 605

```

```

RESULT 31
0981X2 PRELIMINARY; PRT; 551 AA.
ID 0981X2;
AC 0981X2;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein ml12208.
GN ML2208.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxId=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099; PubMed=11214968;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Ref. 7:331-338(2000).
DR EMBL; AP002999; BAB49394.1;
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 11.
DR SMART; SM00028; TPR; 11.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 551 AA; 60602 MW; 528D37C365CCFE75 CRC64;

Query Match 10.4%; Score 108; DB 16; Length 551;
Best Local Similarity 25.5%; Pred. No. 0.6;
Matches 47; Conservative 32; Mismatches 69; Indels 36; Gaps 7;

QY 30 NPOLAQIRTOIAISLDMGKLDQAQQLDALSADRFAPAYRTLKYQASEDATHQ 87
DB 91 NPELSAAYNNNGILVLVAKGYDRAITLDDAIFLDPDPAFFYNNRGAASYKAND----- 146
QY 88 TKQQLFEKAIELNPKMQSYMDGYFLVQMGDLSGALLYFDKPSRAIYEGRVAIE-N 146
DB 147 -RAIADYDAIKLNPGBARAYHNHNGLNMARKD-----KERALIDYSAISLDPR 195
QY 147 MAYIYHOYEAAKSPKDDYNNKSALEBALISGTHODEIKSYDKL 193
DB 196 NASSYNNRGAWSKGDDBRAMADYNOVITLDTKNAHAAYRRGLIWSRKDD-----SRA 250
QY 194 LSDY 197
DB 251 IADY 254

```

RESULT 32
Q8YD43

ID Q8YD43 PRELIMINARY; PRT; 255 AA.
 AC Q8YD43;
 RA 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 GN Tetrahymena ciliate repeat family protein.
 OS BMEI10334.
 OC Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 NC NCB1_TaxID=29459;
 RX NCB1_TaxID=29459;
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=1175668;
 RA Delvecchio V.G., Kapareit V., Redkar R.J., Patra G., Mujar C., Los T.,
 Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 Jadhav L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldman E.,
 Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Teleson J.-J.,
 RA Hasekorn R., Kyriakides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis." RT
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AB009671; AAL53576.1; -.
 DR InterPro; IPR00504; RNA_rec_mot.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 6.
 DR SMART; SMO0028; TPR; 6.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 255 AA; 28059 MW; 17CE1CC26B63C0C9 CRC64;

Query Match 10.3%; Score 107.5; DB 16; Length 255;
 Best Local Similarity 29.9%; Pred. No. 0.24; Matches 41; Conservative 14; Mismatches 67; Indels 15; Gaps 5;

QY 48 GKLDAKQOALDALSADROFAPAYRTLAKVYQASDATHQTKAQLFEKALIELNPKMQS 107
 DB 122 GHLDPALNDPQALALRTTDSRAVHNRGLIYQAK--GLHK-QAIDFSKATSLNSTAPR 178
 QY 108 YMDYGFYLVOMGDLGSLALYFPKPSRAIGYGRV-AIENNAVITYHOREAKSPFK-- 163
 DB 179 YNGRISYVALDQDYNA---FDDFNTATLTLDQNVASWAMQALVYEHNGDKAKANSYAR 235
 QY 164 -----DQYNAKSALEP 175
 DB 236 AVQDLPKTKPKAGGLAR 252

RESULT 33
 Q8TR22 PRELIMINARY; PRT; 400 AA.
 ID Q8TR22;
 AC Q8TR22;
 RA 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 GN O-linked G1CNAc transferase.
 OS MA1362.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NC NCB1_TaxID=2214;
 RX NCB1_TaxID=2214;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smitrov S., Alnoor D., Brown A.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Titrrell A., Ye W.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,

RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity." RT
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE010806; AAM04778.1; -.
 KW Transferase; Complete proteome.
 SQ SEQUENCE 400 AA; 45218 MW; BAA79A182FA20607 CRC64;

Query Match 10.3%; Score 107.5; DB 17; Length 400;
 Best Local Similarity 23.7%; Pred. No. 0.43; Matches 45; Conservative 30; Mismatches 70; Indels 45; Gaps 5;

QY 29 NNPOAQRRTQIAISLDMGKLDQAKQOALDALSADROFAPAYRTLAKVY-----QASE 82
 DB 48 SNEIAMELNCGDLKLGKFNALIAFDKAIKDPK-----NIYLLNKKAALE 98
 QY 83 DATQTKAQLFEKALIELNPKMQSYMDYGFYLVOMGDLGSLALYFPK----- 131
 DB 99 SFGFEPLKLYQAVELNSDADLNNMAFSYQGEYERAVAYGKALELRDYPNAW 158
 QY 132 -----SRAIGYGRVVAIENNAVITYHOREAKSPFKDQYNAKSALEPALLISGTOHD 184
 DB 159 YGKALNLSQAGSYEAEVAYEKV-----LEESP---DYKAMMGKGIALQMGWRYD 206
 QY 185 EIKSYDKL 194
 DB 207 EAIAYDKAI 216

RESULT 34
 Q8TR22 PRELIMINARY; PRT; 914 AA.
 ID Q8TR22;
 AC Q8TR22;
 RA 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 GN Hypothetical protein MA0188.
 OS MA0188.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NC NCB1_TaxID=2214;
 RX NCB1_TaxID=2214;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smitrov S., Alnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Titrrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity." RT
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE010676; AAM03641.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 914 AA; 106181 MW; CF408FFB289B7F64 CRC64;

Query Match 10.3%; Score 107; DB 17; Length 914;
 Best Local Similarity 23.5%; Pred. No. 1.4; Matches 54; Conservative 40; Mismatches 80; Indels 56; Gaps 12;
 QY 8 PWNAMGIVLSACOSTRIPKNNPQALQIRTOIAISILD-----MGKLDQAKQOALDALS 62

DB 496 PWYEMTLQILEA-----KHGPERKDIVT--TNIILDLVYKNGEYKALQSERALA 545
 QY 63 ABR-----OPAPAYRTLAVYQASSEDATQHTQAORLFEKALIELNK-----DMOS 107
 DB 546 IGETTLGVQHPDVATSLDNLGLY---ESMGNYKQALQSERALIEYKVLGPQHRDVAI 602
 QY 108 YMD--YGFVLVOMGDLGSLIYEDK---PSRAIG--YEGRVVAIENNAVYHYHQAIAK 159
 DB 603 TLIDNLGLY--ESMGYERKALITVORTTIKEKVLGPQHSNPAISLIDNLAVLYRQWGE--- 658
 QY 160 SPTKDDYNNAKSALERAL-----ISGTHDEIKSYDILSLDYKLSPDK 204
 DB 659 -----YKALQLSQRALEIYKVLGPQHPDIATLNNIALLYOSMEDYQ 702

RESULT 35
 ID Q92BK3 PRELIMINARY; PRT; 212 AA.
 AC Q92BK3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein lin1345.
 GN LIN1345.
 OS Bacteria.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 OX NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / SEROVAR 6A;
 RX PubMed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Ruenick C., Amend A.,
 Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
 Charbit A., Cherouni F., Couve E., de Darvar A., Denoux P.,
 Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauguet O.,
 Entian K.-D., Fish H., Garcia-del Portillo F., Garrido P.,
 Gautier L.-M., Kaerst U., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 Jones L. M., Kaefer U., Krefit J., Kuhn M., Kunst F., Kuruphat G.,
 Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,
 Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 Remmel B., Rose M., Schluter T., Simoes N., Tietze A.,
 Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.,
 RT "Comparative genomics of Listeria species."
 RL Science 294:849-852(2001).
 DR EMBL; AL596168; CAC96776.1; --
 DR Listeriol; LIN01545;
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 5.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 212 AA; 24098 MW; 015D8C89534E213B CRC64;

Query Match 10.2%; Score 106.5; DB 16; Length 212;
 Best Local Similarity 25.0%; Pred. No. 0.23;
 Matches 43; Conservative 22; Mismatches 66; Indels 41; Gaps 4;
 QY 44 LIDMGKLDQAKQQLDAALSADROFAPAYRTLAVY-----QASDATHQ----- 87
 DB 34 LISMDFERALELPKALIEDLDTVPAYAYSLGLIYELRYGEADSPFNATKQGMENGND 93
 QY 88 -----TKAORLEFKAELNPKMOSYMDGFLVOMGDLGSLIYFDKPS 132
 DB 94 LFFMLGMSFYQMEELTLAMPYLLRSVELNPDESEALFOYGIVLASG-----FYED--- 144
 QY 133 RAIGYGRVVAIENNAVYHYHQAIAKSPKDDYNNAKSALERALISGTQHD 184
 DB 145 -AIIIMLERYLLIKQEDPALVNIIGAAYLWQGDIVLAKNYFERALATGIPHE 195

RESULT 36
 ID Q51526 PRELIMINARY; PRT; 198 AA.
 AC Q51526;

DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE p11f protein.
 GN p11f.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAOI;
 RX MEDLINE=97128806; PubMed=8973346;
 RA Watson A.A., Alm R.A., Mattick J.S.;
 RT "Identification of a gene, p11f, required for type 4 fimbrial
 biogenesis and twitching motility in Pseudomonas aeruginosa."
 RL Gene 180:49-56(1996).
 DR EMBL; L49434; AA840949.1; --
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 3.
 SQ SEQUENCE 198 AA; 22416 MW; 6B1DE4659D2D408E CRC64;

Query Match 10.1%; Score 105; DB 2; Length 198;
 Best Local Similarity 24.5%; Pred. No. 0.28;
 Matches 47; Conservative 26; Mismatches 81; Indels 38; Gaps 5;
 QY 1 WKIRYKMWVAMGL-----VLSACQSTPIPPKNNPQLAQRTOAISLDMGKLDQAKQQL 57
 DB 1 MTRPAALVELLAVGLGCVTSGDQNPDKTKGREGADATVQLGTLQNGNTQAKVPL 60
 QY 58 DAALSADROFAPAYRTLAVYQAS-----EDAT 85
 DB 61 RKALIEPSSADAAALAAVAFQTEMEPKADEBYRKALASDNNARYLNNYGGFLYEQR 120
 QY 86 HQTKAORLEFKAELNPKMOSYMDGFLVOMGDLGSLIYEDKPSRAIGYGRVVAI 144
 DB 121 YEAYQGLLEMSQDTLYPERSRPFENIGLVSLQMKKPAQAKYFEKSLR-LNNQPSVAL 179
 QY 145 ENNAVYHYHQA 156
 DB 180 E-MADLYKERE 190

RESULT 37
 ID Q8YOP7 PRELIMINARY; PRT; 707 AA.
 AC Q8YOP7;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Serine/threonine kinase.
 GN ALI3773.
 OS Ambaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 Watanabe A., Itiguchi M., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Ambaena sp. strain PCC 7120."
 RL DNA Res. 8:205-213(2001).
 DR EMBL; AP003594; BAB75472.1; --
 DR InterPro; IPR000719; Btk_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR InterPro; IPR001440; Ser_thr_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; PKinase; I.
 DR Pfam; PF00515; TPR; 10.

DR SMART; SM00220; S_TKC; 1.
 DR SMART; SM00228; TPR; 10.
 DR SMART; SM00219; TYRC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 707 AA; 80375 MW; 2D1B540ADE551CF CRC64;

Query Match 10.0%; Score 104.5; DB 16; Length 707;
 Best Local Similarity 22.8%; Pred. No. 1.7;
 Matches 50; Conservative 36; Mismatches 72; Indels 61; Gaps 8;

QY 36 IFTQAIISLDGKLDQAKQDLAALSADRO-----PAPY-----FTLAKVQASEDA 84
 DB 324 INSNNTADLAKKGNFPELQRYKDALSAVEQAVDIRPDYAPAMQGGKTLFRLKQYQDAL 383
 QY 85 THTQAKQRLFEKAIELNPKDMQSYNDYGYLVQMGDSLALYFPKP----- 131
 DB 384 T-----AYDKAIQIQDPYVAMSGRGFSLQNLQRYSEALISFPKALQNLNENYEPVWNA 436
 QY 132 -----SRAIGEGRVVAIENMAY-IYHYQVEAKS-----PTKD 164
 DB 437 REEASNLKQYDQALIKSYDKALIEPNSDAESFYNGKGLQSKMEINENAYNKAIEIKS 496
 QY 165 DYNNAKSALERALISGTHDEIKSYDKLSYDKLSYD 203
 DB 497 DYERAWYNLGNLVLNRYEDAFKAYDKAV-QYK--TDY 532

RESULT 38

QY 08R147 PRELIMINARY; PRT; 628 AA.
 AC 08R147;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Tetratricopeptide repeat family protein.
 GN FN1787.
 OS Fusbacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusbacteria; Fusbacterium.
 NC NCB1_Taxid=76856;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapreth V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Golsman E., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 RA Fontein M., Kyriades N., Overbeek R.,
 RT "Genome sequence and analysis of the oral bacterium Fusbacterium
 nucleatum strain ATCC 25586".
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL; AB010481; AAL93886.1; -.
 KW Complete proteome.
 SQ SEQUENCE 628 AA; 73857 MW; 013321316E4829B4 CRC64;

Query Match 10.0%; Score 104; DB 16; Length 628;
 Best Local Similarity 26.0%; Pred. No. 1.6;
 Matches 45; Conservative 27; Mismatches 65; Indels 36; Gaps 8;

QY 41 AISLDMGKLDQAKQDLAALSADROFA-----PAYTTLKVVQASDATHQTKAQL 93
 DB 68 ACSYCSNKKYDKALIEDYDKALINPDACYFNNRGHSYALKYSEALD----- 117
 QY 94 FEKAIELNPKDMQSYNDYGYLVQMGDSLALYFPK-----PSRAIGEGRVVAIENMA 148
 DB 118 YDKAIKLPNNASYYKGGFSYALNKYDKALIEDYKAIKLPNNAAVPSR-----G 170
 QY 149 YVYHYQVEAKSPYDQVWNA-KSALERALI-----SGTHDEIKSYDKLSY 197
 DB 171 DIYI--YEKAVKSIEDYKAIKLPNNAFYDNRGLAYEKLK--YKEAINDY 220

RESULT 39

QY 08TNN0 PRELIMINARY; PRT; 466 AA.
 AC 08TNN0;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Kinesin light chain.
 GN MA2255.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NC NCB1_Taxid=2214;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
 RA Allen N., Naylor J., Strange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talmas J., Tirrell D.A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grathame J.A.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
 RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.,
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AB010912; AAM05648.1; -.
 KW Complete proteome.
 SQ SEQUENCE 466 AA; 53703 MW; CAF69337C442687 CRC64;

Query Match 9.9%; Score 103.5; DB 17; Length 466;
 Best Local Similarity 22.4%; Pred. No. 1.2;

Matches 46; Conservative 32; Mismatches 80; Indels 47; Gaps 7;

QY 10 VMANGVLISACQSPPIPKNNPOLAQRTOAIISLDGKLDQAKQDLAALSADR----- 65
 DB 213 IRALGLEKLEKEPEENKVALAYTAGTLNNGVLISEGKEBAERVGALKIQEKIYG 272
 QY 66 -----QFAPAYTTLKVVQASDATHQTKAQLFEKAIELNPKDMQSYNDYGYLVQMGDL 121
 DB 273 NEHPQIAQTLNLLLY--FQTRYEKAMLYTRSLIEMK-----LGKT 315
 QY 122 SCALYFDPKPSRAIGEGRVVAIENMA--YVYHYQVEAKSPYDQVWNAKSALERALIS 179
 DB 316 EHA-----GFATTNNAGVYQKRYEKA-----LELYTRALEIERIL-- 355
 QY 180 GTQHDEIKSYDKLSYDKLSYD 204
 DB 356 GPDNPEVAKTLNNGELVRLIGQHK 380

RESULT 40

QY 082422 PRELIMINARY; PRT; 944 AA.
 AC 082422;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Gibberellin action negative regulator spy.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Hordeum.
 NC NCB1_Taxid=4513;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. HIMALAYA;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:34:10 ; Search time 70 Seconds
(without alignments)
388.330 Million cell updates/sec

Title: US-10-069-544-2

Perfect score: 204

Sequence: 1 MKIRYKMPWVAMGLVLSAC.....EIKSYDKLLSDYKLLSDYK 204

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A.Geneseq_101002.*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	204	100.0	204	22	AAAB60638
2	11	5.4	327	22	AAAB11770
3	10	4.9	122	21	AAAB29899
4	8	3.9	212	21	AAAG19727
5	8	3.9	212	21	AAAG32512
6	8	3.9	212	21	AAAG45460
7	8	3.9	212	21	AAAG50087
8	8	3.9	232	21	AAAG45459
9	8	3.9	232	21	AAAG50086
10	8	3.9	1120	20	AAAB16462

11	8	3.9	2154	20	AAAB1639
12	7	3.4	34	23	AAU78155
13	7	3.4	93	22	AAAB9285
14	7	3.4	110	21	AAAG3742
15	7	3.4	122	23	AAAG80155
16	7	3.4	140	21	AAAG19728
17	7	3.4	140	21	AAAG32513
18	7	3.4	140	21	AAAG45461
19	7	3.4	140	21	AAAG50088
20	7	3.4	143	18	AAAB28115
21	7	3.4	178	23	ABP28232
22	7	3.4	215	21	AAV75856
23	7	3.4	260	22	ABG09899
24	7	3.4	275	19	AAV11104
25	7	3.4	285	19	AAV11105
26	7	3.4	288	19	ABP25671
27	7	3.4	293	21	AAAB98446
28	7	3.4	293	21	AAAG40720
29	7	3.4	330	20	AAAB88339
30	7	3.4	333	21	AAAG40719
31	7	3.4	336	21	AAV95047
32	7	3.4	349	23	AAU69429
33	7	3.4	350	22	ABG18243
34	7	3.4	358	21	AAAG40718
35	7	3.4	360	22	ABAB60104
36	7	3.4	378	22	AAAB88515
37	7	3.4	378	23	AAU91446
38	7	3.4	382	23	AAU77092
39	7	3.4	444	22	ABG30309
40	7	3.4	477	22	ABG22445
41	7	3.4	490	22	ABG07778
42	7	3.4	490	22	ABG08484
43	7	3.4	582	21	AAV93405
44	7	3.4	592	22	ABG35607
45	7	3.4	749	22	ABG05779
46	7	3.4	1017	23	AAU81761
47	7	3.4	1095	22	AAAG83030
48	7	3.4	1272	22	ABG03108
49	7	3.4	10182	23	ABP38314
50	6	2.9	9	22	AAAB83907
51	6	2.9	15	18	AAAB8940
52	6	2.9	21	22	AAAB6047
53	6	2.9	22	13	AAAB26542
54	6	2.9	22	22	AAAB87048
55	6	2.9	26	22	ABB39281
56	6	2.9	26	22	AAAB5947
57	6	2.9	26	22	AAW72546
58	6	2.9	26	22	AAAB32786
59	6	2.9	26	23	ABG42366
60	6	2.9	31	16	AAAB85039
61	6	2.9	32	23	AAU69662
62	6	2.9	33	23	ABP77864
63	6	2.9	34	23	AAU78158
64	6	2.9	45	21	AAV91501
65	6	2.9	48	21	AAV64661
66	6	2.9	49	22	AAU00880
67	6	2.9	54	22	AAU67516
68	6	2.9	54	22	AAAB89014
69	6	2.9	55	21	AAAG02418
70	6	2.9	62	21	AAV97115
71	6	2.9	62	22	AAU49884
72	6	2.9	66	22	AAU78640
73	6	2.9	66	22	ABG08737
74	6	2.9	66	23	ABP26077
75	6	2.9	67	22	AAU1974
76	6	2.9	68	22	ABB38525
77	6	2.9	68	22	AAAB59144
78	6	2.9	68	22	AAW71680
79	6	2.9	68	22	AAAB1973
80	6	2.9	68	23	ABG41493
81	6	2.9	69	21	AAAG36422
82	6	2.9	70	22	AAU56487
83	6	2.9	70	22	AAW79624

Mouse elf-1 protei
HIV TetraTicopept
Human immune/haema
Human secreted pro
SCT domain protein
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Streptococcus aur
Streptococcus poly
Neisseria strain g
Novel human diagno
H. pylori ORF 05ce
H. pylori ORF hp7e
Streptococcus poly
H. pylori GHP0_542
Zea mays protein f
Salmonella enteric
Zea mays protein f
Candida albicans p
Lung small cell ca
Novel human diagno
Zea mays protein f
Drosophila melanog
Haemophilus influe
Haemophilus influe
Human viral protei
Novel human diagno
Novel human diagno
Novel human diagno
Novel human diagno
Amino acid sequenc
Novel human diagno
Novel human diagno
S. ephelantiidis ope
A. thaliana putati
Novel human diagno
Staphylococcus epi
Amino acid sequenc
Peptide resembli
Human TANGO 265 tr
Sequence of endopr
Human TANGO 339 tr
Peptide #6787 enco
Human brain expres
Human bone marrow
Human bone marrow
Peptide #6823 enco
Human peptide enco
Peptide RI from th
Cell death protect
Phosphatase 1 bind
Human TetraTicope
Human secreted pro
Human 5' EST relat
Human cancer relat
Human immune/haema
Human secreted pro
Tau conotoxin prot
Propionibacterium
Human protein SEQ
Propionibacterium
Novel human diagno
Streptococcus poly
Novel human respit
Peptide #6031 enco
Human brain expres
Human bone marrow
Peptide #6010 enco
Human peptide enco
Arabidopsis thalia
Propionibacterium
Human protein SEQ

W 3 Jul 9 09:59:16 2003

230	6	2.9	191	21	AAI5550	Human immune syste
231	6	2.9	191	21	AAV74969	Neisseria gonorrhoe
232	6	2.9	191	21	AAV74970	Neisseria meningit
233	6	2.9	191	22	ABB67051	Drosophila melanog
234	6	2.9	192	21	AAI56077	Human secreted pro
235	6	2.9	193	21	AAI23443	Arabidopsis thalia
236	6	2.9	193	22	AAI87037	Human TANGO 339 tr
237	6	2.9	194	22	AAU51709	Propionibacterium
238	6	2.9	197	21	AAI90334	B. subtilis nitrore
239	6	2.9	199	22	AAI81852	S. epidermidis ope
240	6	2.9	200	22	AAI66497	Drosophila melanog
241	6	2.9	200	22	AAI66635	Drosophila melanog
242	6	2.9	200	22	AAI79781	Corynebacterium gl
243	6	2.9	203	21	AAI21974	Arabidopsis thalia
244	6	2.9	203	22	AAI70414	Drosophila melanog
245	6	2.9	206	21	AAI54538	Zea mays protein f
246	6	2.9	208	22	AAI61552	Novel human diagno
247	6	2.9	208	22	AAI92559	Human protein sequ
248	6	2.9	209	21	AAI23047	Human tetraspanin-
249	6	2.9	209	22	AAU37810	Streptococcus pneu
250	6	2.9	210	21	AAI606898	Arabidopsis thalia
251	6	2.9	210	23	AAI39945	Staphylococcus epi
252	6	2.9	210	23	AAU74389	Breast tumour-spec
253	6	2.9	211	22	AAI80952	Human nPCR36 #2.
254	6	2.9	213	21	AAI2120	Hydrophobic domain
255	6	2.9	213	21	AAI43037	Human ORF2801
256	6	2.9	213	21	AAI26619	Arabidopsis thalia
257	6	2.9	213	21	AAI48629	Arabidopsis thalia
258	6	2.9	213	23	AAI27766	Streptococcus poly
259	6	2.9	213	23	AAI30083	Streptococcus poly
260	6	2.9	214	22	AAI04557	Novel human diagno
261	6	2.9	216	22	AAI91994	C glutamicum prote
262	6	2.9	216	23	AAI55277	Listeria monocytog
263	6	2.9	216	23	AAI49513	Listeria monocytog
264	6	2.9	217	22	AAI80930	Human nPCR3 #1.
265	6	2.9	218	20	AAI36106	Extended human sec
266	6	2.9	219	13	AAI26540	Sequence of 40 kD
267	6	2.9	219	22	AAU87297	Novel central nerv
268	6	2.9	220	20	AAI37376	Protein involved i
269	6	2.9	220	21	AAI6474	Arabidopsis thalia
270	6	2.9	220	21	AAI39035	Arabidopsis thalia
271	6	2.9	220	22	AAI93018	C glutamicum prote
272	6	2.9	221	22	AAI81049	Human protein HP00
273	6	2.9	222	19	AAI60959	Streptococcus pneu
274	6	2.9	224	20	AAI56764	Protein involved i
275	6	2.9	224	21	AAI05764	Arabidopsis thalia
276	6	2.9	224	21	AAI23442	Arabidopsis thalia
277	6	2.9	226	21	AAI94879	Human protein clon
278	6	2.9	226	22	AAI39575	Human polypeptide
279	6	2.9	226	22	AAI74720	Human membrane ass
280	6	2.9	226	22	AAI88317	Human membrane or
281	6	2.9	226	23	AAI89618	Human polypeptide
282	6	2.9	228	22	AAI6581	Human novel cytoxi
283	6	2.9	228	22	AAI87035	Human mature TANGO
284	6	2.9	230	22	AAI2898	Novel human secret
285	6	2.9	230	23	AAI28161	Streptococcus poly
286	6	2.9	231	16	AAI66292	Mycobacterium bovi
287	6	2.9	231	16	AAI63900	M. bovis InhA. My
288	6	2.9	231	16	AAI63900	Drosophila melanog
289	6	2.9	231	22	AAI4361	Human polypeptide
290	6	2.9	232	22	AAI91829	C glutamicum prote
291	6	2.9	232	22	AAI79921	Corynebacterium gl
292	6	2.9	232	22	AAI9529	Arabidopsis thalia
293	6	2.9	234	21	AAI06055	Arabidopsis thalia
294	6	2.9	234	21	AAI53067	Arabidopsis thalia
295	6	2.9	234	22	AAI82995	Arabidopsis thalia
296	6	2.9	236	20	AAI31829	Human adult blood
297	6	2.9	236	20	AAI36004	Extended human sec
298	6	2.9	236	20	AAI02619	Amino acid sequenc
299	6	2.9	236	21	AAI06897	Arabidopsis thalia
300	6	2.9	236	21	AAI11044	Arabidopsis thalia
301	6	2.9	236	21	AAI44362	Human cell cycle r
302	6	2.9	236	22	AAI93724	Human polypeptide,

668	2.9	443	20	AAV55551	Nematode STE20-rel
669	2.9	443	21	ABP93856	Herbicidically activ
670	2.9	444	21	AAAG36353	Arabidopsis thalia
671	2.9	446	21	AAAR15499	Rat dopamin D1 re
672	2.9	446	21	AAAG39348	Arabidopsis thalia
673	2.9	446	22	AAAB76810	Corynebacterium gl
674	2.9	447	22	ABG08287	Novel human diagn
675	2.9	448	22	AAAB25874	Human protein sequ
676	2.9	451	23	AAAG91304	C glutamicum prote
677	2.9	451	23	ABP39575	Staphylococcus epi
678	2.9	452	22	ABG16289	Novel human diagn
679	2.9	453	22	AAAM39242	Human polypeptide
680	2.9	453	22	AAAB66833	Human GTP-binding
681	2.9	454	22	AAU33381	Enterococcus faeca
682	2.9	454	22	ABG25565	Novel human diagn
683	2.9	455	21	AAAG48082	Arabidopsis thalia
684	2.9	455	22	AAU38201	Salmonella typhi c
685	2.9	459	22	ABG29005	Novel human diagn
686	2.9	459	22	ABG29867	Novel human diagn
687	2.9	460	22	AAU33561	Pseudomonas aerugi
688	2.9	463	20	AAAB89250	Mouse P705 Isotof
689	2.9	463	21	AAAG42251	Arabidopsis thalia
690	2.9	466	22	AAAB95611	Human protein sequ
691	2.9	466	22	ABP26809	Streptococcus poly
692	2.9	467	22	AAAG91066	C glutamicum prote
693	2.9	468	22	AAAB83855	Amino acid sequenc
694	2.9	469	22	AAAB60093	Putative P. adyresi
695	2.9	469	23	ABP90252	Human polypeptide
696	2.9	470	23	ABP72350	Murine protein iso
697	2.9	471	13	AAAR27246	Sequence encoded b
698	2.9	471	13	AAAY7689	GRav-5 ORF2 prote
699	2.9	473	21	AAAG42250	Arabidopsis thalia
700	2.9	474	22	ABBS58675	Drosophila melanog
701	2.9	474	22	AAU35180	Enterococcus faeca
702	2.9	474	22	AAU35180	Human lipid metabo
703	2.9	475	21	ABAB07495	Human cancer assoc
704	2.9	475	21	AAAY74971	Neisseria meningit
705	2.9	475	21	AAAG9363	C glutamicum prote
706	2.9	481	21	AAAG42581	Arabidopsis thalia
707	2.9	481	21	ABBS59486	Arabidopsis thalia
708	2.9	482	21	AAAG30065	Arabidopsis thalia
709	2.9	482	21	AAAG42249	Arabidopsis thalia
710	2.9	483	21	AAAG36884	Arabidopsis thalia
711	2.9	483	21	AAAG36352	Arabidopsis thalia
712	2.9	484	8	AAAP71081	Sequence encoded b
713	2.9	484	21	AAAG35683	Arabidopsis thalia
714	2.9	485	21	AAAG35682	Arabidopsis thalia
715	2.9	487	12	AAAR1596	D1 dopamine recept
716	2.9	487	18	AAAB09795	S. epidermidis ope
717	2.9	489	22	AAAG82769	Arabidopsis thalia
718	2.9	491	21	ABBT0480	Drosophila melanog
719	2.9	496	21	AAAG36351	Arabidopsis thalia
720	2.9	496	22	ABBS59503	Drosophila melanog
721	2.9	496	22	ABBS59503	Novel human diagn
722	2.9	496	22	AAAB07503	Arabidopsis thalia
723	2.9	496	23	ABBS92385	Herbicidically activ
724	2.9	497	22	ABBS59501	Drosophila melanog
725	2.9	499	22	AAAB09722	Novel cell cycle p
726	2.9	499	22	AAAB09731	Protein phosphatas
727	2.9	499	22	AAAB09732	Protein phosphatas
728	2.9	499	22	AAAB09733	Protein phosphatas
729	2.9	500	22	ABG09989	Novel human diagn
730	2.9	502	21	AAAG30064	Arabidopsis thalia
731	2.9	502	22	AAU92223	Human P70 polypept
732	2.9	502	22	AAAB07160	Human interleukin-
733	2.9	502	22	AAAB24349	Human EST encoded
734	2.9	502	22	AAU04955	Human interleukin
735	2.9	502	22	AAAB06586	Human protein havi
736	2.9	502	22	AAAB7604	Human PRO5801. HO
737	2.9	502	22	AAAB18127	Human DNAX cytokin
738	2.9	502	23	AAU11351	Human DNAX cytokin

KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antistatic; antirheumatic; haemostatic; antitumor; antileukemic;
KW cytotoxic; osteoporosis; vasodilator; cardiac; virucide; antibacterial;
KW antifungal; vulnery; antitumor.
XX
OS Homo sapiens.
XX
PN MO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001MO-US03800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PA (HYSB-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR MPI; 2001-457740/49.
DR N-PSDB; ABA09014.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
XX
PS Claim 20; Page 243; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoietic regulatory activity; tissue growth activity;
CC immunomodulatory activity; or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
SQ Sequence 327 AA;
XX

Query Match 5.4%; Score 11; DB 22; Length 327;

Best Local Similarity 100.0%; Pred. No. 0.0093;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 FEKATLEINPKD 104
Db 205 FEKATLEINPKD 215

RESULT 3

AA29889
ID AAB29889 standard; Protein; 122 AA.

AC AAB29889;
XX

DT 09-FEB-2001 (first entry)
XX

DE Human secreted protein BLAST search protein SEQ ID NO: 157.
XX

XX Cytostatic; immunosuppressive; nocitropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX

OS Homo sapiens.
XX

PN MO200061779-A1.
XX

PD 19-OCT-2000.
XX

PF 06-APR-2000; 2000MO-US09068.
XX

PR 09-APR-1999; 99US-0126699.
XX

PR 20-JUN-2000; 2000US-0177050.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX

PI Rosen CA, Ruben SM, Komatsoulis G;
XX

DR MPI; 2000-647424/62.
XX

PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX

PS Disclosure; Page 478-479; 495pp; English.
XX

XX The invention relates to the isolation of genes AC63410-C63458 encoding
CC the human secreted proteins AAB29889-829850. This sequence represents a
CC peptide fragment homologous to the protein encoded by the gene given
CC in the descriptor line. The sequence is a search result from a BLASTX
CC homology search. The genes and proteins are useful for preventing,
CC ameliorating or treating medical conditions, e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such
CC as myocardial ischaemia; (d) wound healing; (e) neurological diseases
CC e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such
CC as viral, bacterial, fungal and parasitic infections.
XX

SQ Sequence 122 AA;
XX

Query Match 4.9%; Score 10; DB 21; Length 122;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 FEKATLEINPK 103
|||||

Db 9 FEKAELNPK 18

RESULT 4

AAG19727

ID AAG19727 standard; Protein: 212 AA.

XX AAG19727;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21637.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EPI033405-A2.

PD

XX

XX 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0131825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 26-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132486.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 14-JUN-1999; 99US-0138847.

PR 16-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 18-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 24-JUN-1999; 99US-0140354.

PR 28-JUN-1999; 99US-0140695.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0142154.

PR 01-JUL-1999; 99US-0142187.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143524.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144684.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145513.

PR 27-JUL-1999; 99US-0145818.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147933.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155186.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.9%; Score 8; DB 21; Length 212;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AMGLVLSA 19
|||
DB 72 AMGLVLSA 79

RESULT 5
AAG32512
ID AAG32512 standard; Protein; 212 AA.
XX

AC AAG32512;
XX 17-OCT-2000 (first entry)
DT Arabidopsis thaliana protein fragment SEQ ID NO: 39236.
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 39236.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.

```
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0136264.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147499.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.9%; Score 8; DB 21; Length 212;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12. AMGLVLSA 19
Db 72 AMGLVLSA 79

RESULT 6
AAG45460
ID AAG45460 standard; Protein; 212 AA.
XX
AC AAG45460;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 57077.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
```


PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158028.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 3.9%; Score 8; DB 21; Length 212;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AMGLVLSA 19
Db 72 AMGLVLSA 79

RESULT 7
AAG50087
ID AAG50087 standard; Protein: 212 AA.

XX AAG50087;
DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 63434.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
XX termination sequence.

XX OS Arabidopsis thaliana.

XX EPI033405-A2.

XX PR 06-SEP-2000.
XX PF 25-FEB-2000; 2000EB-0301439.
XX XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123748.
PR 22-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130456.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134421.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142893.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144652.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.9%; Score 8; DB 21; Length 212;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 AMGLVLSA 19
|||||
Db 72 AMGLVLSA 79

RESULT 8
ID AAG45459 standard; Protein; 232 AA.
AAG45459;

AC AAG45459;
DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 57076.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX Arabidopsis thaliana.
XX EPI033405-A2.

XX EPI033405-A2.
XX PD 06-SEP-2000.

XX PD 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.

XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135351.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138099.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 26-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 3.9%; Score 8; DB 21; Length 232;
Matches 8; Conservative 100.0%; Pred No. 8; Mismatches 0; Indels 0; Gaps 0;

Qy 12 AMGLVLSA 19
Db 92 AMGLVLSA 99

RESULT 9
AAG50086
ID AAG50086 standard; Protein; 232 AA.

AC AAG50086;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 63433.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

```

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144432.
PR 20-JUL-1999; 99US-014484.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

```

```

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      3.9%; Score 8; DB 21; Length 232;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      12 AMGLVISA 19
Db      92 AMGLVISA 99

```

```

RESULT 10
AAW81642
ID      AAW81642 standard; Protein; 1120 AA.
XX

```

```

AC      AAW81642;
XX

```

```

DT      01-MAR-1999 (first entry)
XX

```

```

DE      Mouse elf protein.
XX

```

```

KW      Elf; mouse; liver development; signal transduction; biliary tree;
XX

```

```

KW      cholestasis; biliary stone; hepatic obstruction; stricture;
XX

```

```

KW      primary biliary cirrhosis; primary sclerosing cholangitis; therapy.
XX

```

```

OS      Mus sp.
XX

```

```

PN      W09648827-A1.
XX

```

```

PD      05-NOV-1998.
XX

```

```

PR      30-APR-1998; 98WO-US08656.
XX

```

```

PR      30-APR-1997; 97US-0841349.
XX

```

```

PA      (MISH/) MISHRA L.
XX

```

```

PI      Mishra L;
XX

```

```

DR      WPI: 1999-009382/01.
XX

```

```

DR      N-PSDB; AAV64424.
XX

```

```

PT      New isolated early liver development genes - used to develop
PT      products for treating, e.g. liver disease, hepatocellular carcinoma,
PT      degenerative neurological disorders, anaemia, ataxia or
PT      haemochromatosis
XX

```

```

PS      Example 2; Fig 16a; 92PP; English.
XX
CC      This is the amino acid sequence of an elf protein of the foetal
XX

```

CC murine liver, as deduced from an isolated partial cDNA clone (see
 CC AAV64424). The invention provides early developing liver proteins
 CC (see AAW81638-42) and the genes coding for them (see AAV64410-24). The
 CC gene were isolated from embryonic cDNA libraries using subtractive
 CC hybridisation. 3 Elf genes (see AAV64411-13) were identified. The
 CC Elf proteins (see also AAW81639-40) are probably important for the
 CC formation of the biliary tree during early liver development. They
 CC are used in a claimed method to treat cholestasis, biliary stones,
 CC liver obstruction, stricture, primary biliary cirrhosis and primary
 CC sclerosing cholangitis. Early developing liver proteins and primary
 CC nucleic acids can also be used in the diagnosis of liver diseases
 CC and other disorders, including those relating to oncogenesis and
 CC tissue repair.

CC Sequence 1120 AA;

Query Match 3.9%; Score 8; DB 20; Length 1120;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 QTKAQLRF 94
 |||||
 Db 323 QTKAQLRF 330

RESULT 11

AAW81639
 AAW81639 standard; Protein; 2154 AA.

AC AAW81639;

DT 01-MAR-1999 (first entry)

DE Mouse elf-1 protein.

XX Elf-1; mouse; liver development; signal transduction; biliary tree;
 KW cholestasis; biliary stone; hepatic obstruction; stricture;
 KM primary biliary cirrhosis; primary sclerosing cholangitis; therapy.

OS Mus sp.

PN WO9848827-A1.

PD 05-NOV-1998.

PF 30-APR-1998; 98WO-US08656.

PR 30-APR-1997; 97US-0841349.

PA (MISH/) MISHRA L.

PI Mishra L;

DR WPI; 1999-009382/01.

DR N-PSDB; AAV64411.

PT New isolated early liver development genes - used to develop
 PT products for treating, e.g. liver disease, hepatocellular carcinoma,
 PT degenerative neurological disorders, amaemia, ataxia or
 PT haemochromatosis

PS Claim 9; Fig 2a; 92pp; English.

CC This is the amino acid sequence of the elf-1 protein of the
 CC developing murine foetal liver. The invention provides early
 CC developing liver proteins (see AAW81638-42) and the genes coding for
 CC them (see AAV64410-24). The genes were isolated from embryonic cDNA
 CC libraries using subtractive hybridisation. 3 Elf genes (see
 CC AAV64411-13) were identified. The elf proteins (see also AAW81640) are
 CC probably important for the formation of the biliary tree during early
 CC liver development. They are used in a claimed method to treat
 CC cholestasis, biliary stones, liver obstruction, stricture, primary
 CC biliary cirrhosis and primary sclerosing cholangitis. Early

CC developing liver proteins and nucleic acids can also be used in the
 CC diagnosis of liver diseases and other disorders, including those
 CC relating to oncogenesis and tissue repair.

CC Sequence 2154 AA;

Query Match 3.9%; Score 8; DB 20; Length 2154;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 QTKAQLRF 94
 |||||
 Db 1357 QTKAQLRF 1364

RESULT 12

AAU78155
 AAU78155 standard; Protein; 34 AA.

AC AAU78155;

DT 05-JUN-2002 (first entry)

DE HIV Tetrairicopeptide repeat, TPR2.

XX Viral protein U binding protein; Ubp; Vpu; Tetrairicopeptide repeat;
 KW TPR2; human immunodeficiency virus; HIV-1; CD4; Gag; anti-HIV;
 KM viral titre.

OS Human immunodeficiency virus type 1.

PN US2002022221-A1.

PD 21-FEB-2002.

PF 29-APR-1999; 99US-0301978.

PR 30-APR-1998; 98US-083567P.

PA (PANG/) PANGANIBAN A T.

PA (CALL/) CALLAHAN M A.

PA (HAND/) HANDLEY M A.

PI Panganiban AT, Callahan MA, Handley MA;

DR WPI; 2002-267445/31.

PT New preparation of viral protein U (Vpu) binding protein (Ubp), for
 PT identifying modulators of interaction between Vpu and Ubp, and between
 PT Gag and Ubp

PS Disclosure; Page 17; 29pp; English.

CC The invention relates to a preparation of viral protein U (Vpu) binding
 CC protein (Ubp), where Vpu is encoded by human immunodeficiency virus
 CC (HIV)-1 that promotes degradation of virus receptor, CD4, and enhances
 CC release of virus particles from cells. Also included are the gene
 CC sequence and protein sequence of human Ubp, a method of identifying
 CC modulators/inhibitors of the Ubp/Vpu interaction comprising: (a) exposing
 CC Vpu and Ubp together in the presence of a candidate inhibitor/modulator
 CC under conditions where Vpu and Ubp can interact when an inhibitor is not
 CC present; and (b) determining whether Vpu/Ubp interaction occurs, an
 CC inhibitor of Ubp/Vpu interaction or Gag/Ubp interaction, an anti-Ubp
 CC antibody, creating peptide fragments of the human Ubp protein and
 CC detecting members of the Ubp superfamily using probes derived from the
 CC human Ubp cDNA. Ubp is useful in an assay to identify modulators
 CC (preferably inhibitors) of the Ubp/Vpu interaction. Ubp is also useful
 CC for determining whether a candidate compound modulates (preferably,
 CC inhibits) the interaction between Gag and Ubp. Ubp is also useful in an
 CC assay to examine the effect of Ubp on particle release. The Ubp gene is
 CC useful as an excellent candidate for evaluation of anti-HIV therapeutics.
 CC The anti-Ubp antibody is useful for detecting Ubp. Overexpression of Ubp
 CC in virus expressing cells resulted in significant reduction of HIV-1

CC virion release, and thus Ubp plays a role in Ypu-mediated enhancement
CC of particle release. Ubp is also useful for enhancing the release of
CC particles to increase the titre of HIV or vectors or other retrovirus
CC vectors. An in vitro system for detecting interaction between Ypu
CC and Ubp, and Gag and Ubp, allows rapid and straight forward examination
CC of mutants and reagents that affect protein interaction. The
CC present sequence is an HIV Tetrairicopeptide repeat, TPR2. TRP are found
CC in the human Ubp protein.

SQ Sequence 34 AA;

Query Match 3.4%; Score 7; DB 23; Length 34;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 KAIELNP 102
Db 23 KAIELNP 29

RESULT 13
AAM89285
ID AAM89285 standard; Protein; 93 AA.
XX
AC AAM89285;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:16878.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosolatic; gene therapy; vaccine; metastasis.
OS Homo sapiens.
XX
XX W0200157182-A2.
PN
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229349.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244126.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249220.
 PR 17-NOV-2000; 2000US-0249221.
 PR 17-NOV-2000; 2000US-0249222.
 PR 17-NOV-2000; 2000US-0249223.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249225.
 PR 17-NOV-2000; 2000US-0249226.
 PR 17-NOV-2000; 2000US-0249227.
 PR 17-NOV-2000; 2000US-0249228.
 PR 17-NOV-2000; 2000US-0249229.
 PR 17-NOV-2000; 2000US-0249230.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-483426/52.
 DR N-PSDB; AAK62066.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 XX
 PS Claim 11; SEQ ID NO 16878; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SO Sequence 93 AA;

Query Match 3.4%; Score 7; DB 22; Length 93;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 LLSYKYL 199
 DB 87 LLSYKYL 93

RESULT 14
 AAG03742
 ID AAG03742 standard; Protein; 110 AA.

XX
 AC AAG03742;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 7823.

XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX
 OS Homo sapiens.

XX
 PN EP1033401-A2.

XX
 PD 06-SEP-2000.

XX
 PF 21-FEB-2000; 2000EP-0200610.

XX
 PR 26-FEB-1999; 99US-0122487.

XX
 PA (GENET) GENSET.

XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX
 DR WPI; 2000-500381/45.

XX
 DR N-PSDB; AAC03748.

XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX
 PS Claim 13; SEQ ID 7823; 71bp + CD-ROM; English.

XX
 CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX
 SO Sequence 110 AA;

Query Match 3.4%; Score 7; DB 21; Length 110;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AMGLVLS 18
 DB 64 AMGLVLS 70

RESULT 15
 AAG80155
 ID AAG80155 standard; peptide; 122 AA.

XX
 AC AAG80155;

XX
 DT 23-JUN-2002 (first entry)

XX
 DE SGT domain protein fragment.

XX
 KW Hop protein; chaperone protein; TPR motif; tetratricco peptide repeat;

KW three-dimensional structure; 3D; cytoskeletal; virucide; immunosuppressive;
 KM antiinflammatory; tumour treatment; autoimmune disease; SGT;
 KM protein coordinate data.
 OS Unidentified.

Key Location/Qualifiers
 FH Region 1..12 /note= "Helix 1A"
 FT Region 17..29 /note= "Helix 1B"
 FT Region 35..46 /note= "Helix 2A"
 FT Region 50..62 /note= "Helix 2B"
 FT Region 69..81 /note= "Helix 3A"
 FT Region 85..97 /note= "Helix 3B"
 FT Region 105..117 /note= "Helix C"

DE10018335-A1.
 04-OCT-2001.
 13-APR-2000; 2000DE-1018335.
 29-MAR-2000; 2000DE-1015748.
 (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.

Moarefi I, Scheufler C, Hartl U, Brinker A;
 WPI; 2002-000454/01.

New spatial model of polypeptide containing a TPR motif, useful for
 identifying inhibitors of interaction between chaperone and Hop
 proteins, potential therapeutics -

Claim 31; Fig 3; 145pp; German.

This invention describes a novel spatial model of a polypeptide (I)
 having at least one amino acid (aa) sequence (II) of a TPR (tetratricio
 peptide repeat) structural motif of a Hop protein, or a derivative of
 (II). The invention also describes a method for (1) producing a crystal
 containing, per asymmetric unit of its unit cell, at least one (I) and
 optionally another compound; (2) preparing the crystals of (1); (3)
 preparing compounds (III) that bind, as ligands, to a structural region
 of a Hop protein (or structurally related protein) by non-covalent
 interaction with the main and/or side chains of aa that are components
 of the TPR domain; (4) preparing peptides (Ia), having the spatial model
 of (1), that can bind to a chaperone protein (CP) but are non-functional
 as regards adapter function; (5) identifying compounds (A) that inhibit
 interaction between CP and Hop proteins; (6) (A) identified by the method
 of (5); (7) representing the three-dimensional (3D) structure of a
 polypeptide of unknown structure but containing at least one TPR motif
 (or its derivative or complex), using the model of (1); (8) identifying
 compounds (A') that inhibit interaction between CP and polypeptides
 containing at least one TPR motif, or its derivatives; (9) pharmaceutical
 compositions containing (A) or (A'); (10) preparing DNA sequences (IV)
 that encode a sequence that contains at least part of proteins Trp1 to 4,
 TRSP, SGR and KRA0719; (11) preparing amino acid sequences (V) encoded
 by (IV); and (12) preparing pharmaceutical composition containing (V),
 their fragments or derivatives. The products of the invention have
 cytoskeletal, virucide, immunosuppressive and antiinflammatory activity.
 The models are used to identify compounds (A) that inhibit interaction
 between Hop and chaperone proteins, potentially useful for treating
 tumours and virus infections, also as immunosuppressives, for treating
 autoimmune diseases or other inflammatory states or guest versus host
 disease. This sequence represents a SGT domain used to illustrate the
 method of the invention.

SO Sequence 122 AA;
 Query Match 3.4%; Score 7; DB 23; Length 122;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 KATELNP 102
 Db 26 KATELNP 32

RESULT 16
 AAG19728
 ID AAG19728 standard; Protein; 140 AA.

AC AAG19728;
 DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21638.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.

OS Arabidopsis thaliana.

EN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134768.
 PR 20-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 04-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140891.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145376.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.4%; Score 7; DB 21; Length 140;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 13 MGLVISA 19

Db 1 MGLVLSA 7

RESULT 17
AAG32513
ID AAG32513 standard; Protein, 140 AA.

AC AAG32513;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SPQ ID NO: 39237.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 16-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136382.

PR 28-MAY-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140891.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 25-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145518.
PR 27-JUL-1999; 99US-0145519.
PR 28-JUL-1999; 99US-0145591.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152353.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155466.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159684.
PR 18-OCT-1999; 99US-0159854.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161930.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.4%; Score 7; DB 21; Length 140;
Best Match Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 MGLVLSA 19
|||||
Db 1 MGLVLSA 7

RESULT 18
AAG45461
ID AAG45461 standard; Protein; 140 AA.

XX AC AAG45461;
XX 18-OCT-2000 (first entry)
DT Arabidopsis thaliana protein fragment SEQ ID NO: 57078.
XX DE
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131448.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134421.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138847.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.

```
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158932.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.4%; Score 7; DB 21; Length 140;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 MGLVISA 19
Db 1 MGLVISA 7

RESULT 19
AAG50088
ID AAG50088 standard; Protein; 140 AA.
XX
AC AAG50088;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63435.
XX
```

KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135533.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 01-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145916.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

```

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157752.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 3.4%; Score 7; DB 21; Length 140;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 MGLVLSA 19
 DB 1 MGLVLSA 7

RESULT 20
 AAW28115
 ID AAW28115 standard; Protein: 143 AA.

AC AAW28115;
 DT 27-AUG-1998 (first entry)
 DE Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antitense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KM vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KM toxic shock syndrome.

XX Staphylococcus aureus.
 OS
 XX

Key Location/Qualifiers
 MISC-difference 3 /note="residue designated X is not defined in
 the specification"

FN NO9730070-A1.
 PD 21-AUG-1997.
 PF 19-FEB-1997; 97WO-US02318.
 PR 20-FEB-1996; 96US-0011888.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Pract JM, Reichard RW, Rosenberg M, Ward JM;
 XX WPI; 1997-424969/39.
 DR N-PDB; AAT84047.
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 XX aureus infection
 PS Claim 6; Page 476; 989pp; English.
 XX The present sequence represents a Staphylococcus aureus protein of
 CC unknown function. The DNA sequence was isolated from a library of
 CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
 CC be used in the construction of ribozymes and antisense sequences to
 CC control the expression of Staphylococcal genes. The DNA sequence is
 CC also useful as a source of regulatory elements for the control of
 CC bacterial gene expression. The present protein may be used to produce
 CC vaccines to enable a host to produce specific antibodies with
 CC antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by S. aureus, and conditions relating to
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.

QY 53 AKQQLDA 59
 DB 62 AKQQLDA 68

Query Match 3.4%; Score 7; DB 18; Length 143;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 21
 ABP28232
 ID ABP28232 standard; Protein: 178 AA.

AC ABP28232;
 DT 02-UTL-2002 (first entry)
 DE Streptococcus polypeptide SEQ ID NO 5640.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KM antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.
 XX WO200234771-A2.
 PN 02-MAY-2002.
 PD 29-OCT-2001; 2001WO-GB04789.
 PF
 XX

PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignani V, Margart Ros Yi, Grandi G, Frazer C;
 PI Tettelin H;
 XX
 DR WPI; 2002-352536/38.
 DR N-PSDB; ABN68863.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 PS Claim 1; Page 3726; 4525BP; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 SQ Sequence 178 AA;
 XX
 XX
 Query Match 3.4%; Score 7; DB 23; Length 178;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 191 DKLSYD 197
 |||||
 Db 118 DKLSYD 124
 |||||
 RESULT 22
 ID AAY75856 standard; Protein; 215 AA.
 AC AAY75856;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria strain gnmzq09 ORF 235 protein SEQ ID NO:1157.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria sp.
 XX
 PN WO957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PD 30-APR-1999; 99MO-US09346.
 PF
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR-) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Frazer C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 PI Petersen J, Piza M, Rappoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PS Example 12; Page 124; 1453BP; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 SQ Sequence 215 AA;
 XX
 XX
 Query Match 3.4%; Score 7; DB 21; Length 215;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 LVLSAQ 21
 |||||
 Db 12 LVLSAQ 18
 |||||
 RESULT 23
 ID ABG09899 standard; Protein; 260 AA.
 AC ABG09899;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #9890.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PD 30-MAR-2001; 2001WO-US08631.
 PF
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX

DR WPI: 2001-639362/73.
 DR N-PSDB; AAS74086.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 20; SEQ ID No 40258; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 SQ Sequence 260 AA;
 QY 58 DAALSAD 64
 DB 228 DAALSAD 234
 Query Match 3.4%; Score 7; DB 22; Length 260;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 24
 ID AAY11104 standard; Protein; 275 AA.
 AC AAY11104;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF 05cel0910_23712780_cl_4 cytoplasmic protein.
 KM Vaccine; probe; diagnostic; ORF; cell envelope protein;
 KM secreted protein; cytoplasmic protein; cellular protein.
 OS Helicobacter pylori.
 PN WO9824475-A1.
 PD 11-JUN-1998.
 PF 05-DEC-1997; 97WO-US22104.
 PR 14-JUL-1997; 97US-0891928.
 PR 05-DEC-1996; 96US-0759625.
 PR 25-MAR-1997; 97US-0823745.
 PS (ASTR) ASTRA AB.
 PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
 WPI; 1998-333051/29.

DR N-PSDB; AAX30633.
 PT New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 PS Claims 37, 41; Page 270-271; 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SQ Sequence 275 AA;
 QY 184 DEIKKSY 190
 DB 18 DEIKKSY 24
 Query Match 3.4%; Score 7; DB 19; Length 275;
 Best Local Similarity 100.0%; Pred. No. 98;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 25
 ID AAY11105 standard; Protein; 278 AA.
 AC AAY11105;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF hp7el0192_23712780_f2_5 cytoplasmic protein.
 KM Vaccine; probe; diagnostic; ORF; cell envelope protein;
 KM secreted protein; cytoplasmic protein; cellular protein.
 OS Helicobacter pylori.
 PN WO9824475-A1.
 PD 11-JUN-1998.
 PF 05-DEC-1997; 97WO-US22104.
 PR 14-JUL-1997; 97US-0891928.
 PR 05-DEC-1996; 96US-0759625.
 PR 25-MAR-1997; 97US-0823745.
 PS (ASTR) ASTRA AB.
 PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
 WPI; 1998-333051/29.
 DR N-PSDB; AAX30634.
 PT New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 PS Claims 37, 41; Page 271-272; 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.

CC Vaccines containing the nucleic acids or proteins are claimed, as are
CC probes containing at least 8 nucleotides from the nucleic acid
CC sequences. The vaccines are useful for treating or reducing the risk of
CC H. pylori infections, and the probes can be used diagnostically for
CC detecting the presence of Helicobacter in a sample. The products are
CC also of use in screening for compounds having the ability to interfere
CC with the H. pylori life cycle or to inhibit H. pylori infection.

XX SQ Sequence 278 AA;

Query Match 3.4%; Score 7; DB 19; Length 278;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 DEIKSY 190
|||||

DB 18 DEIKSY 24

RESULT 26

ID ABR25671 standard; Protein; 285 AA.

XX ABR25671;

XX 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 518.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Maignani V, Margalit Ros YI, Grandi G, Fraser C;

XX Tectelin H;

XX WPI; 2002-352536/38.

XX N-PSDB; ABR6302.

XX New Streptococcus protein for the treatment or prevention of infection

XX or disease caused by Streptococcus bacteria, such as meningitis, and

XX for detecting a compound that binds to the protein -

XX Claim 1; Page 3206; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

XX (Streptococcus pyogenes). comprising one of 5483 sequences (S1), given in

XX the specification. The proteins have antibacterial and antiinflammatory

XX activity. (1), nucleic acids encoding (1), ABV6044-ABV71526 and

XX antibodies that bind (1) are used in the manufacture of medicaments for

XX the treatment or prevention of infection or disease caused by

XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX Nucleic acids encoding (1) are used to detect Streptococcus in a

XX biological sample. (1) is used to determine whether a compound binds to

XX (1). A composition comprising (1) or a nucleic acid encoding (1), may be

XX used as a vaccine or diagnostic composition. The disease caused by

XX Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (1) may be used to recombinantly produce (1) and may be
CC used in gene therapy. Antibodies to (1) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

XX SQ Sequence 285 AA;

Query Match 3.4%; Score 7; DB 23; Length 285;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 LERALIS 179
|||||

DB 270 LERALIS 276

RESULT 27

ID AAW98446 standard; Protein; 288 AA.

XX AAW98446;

XX 31-MAR-1999 (first entry)

XX H. pylori GHPO 542 protein.

XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

XX peptic ulcer disease.

XX Helicobacter pylori.

XX WO9843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

XX 01-APR-1997; 97US-0833457.

XX 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX14165.

XX New isolated Helicobacter polynucleotides - used to develop products

XX for the diagnosis, prevention and treatment of Helicobacter

XX infections and gastrointestinal diseases

XX Claim 8; Page 773-774; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHPO protein of the

XX invention. The polypeptides can be used for preventing or treating

XX Helicobacter infections, and gastroduodenal diseases associated with

XX these infections, including acute, chronic, and atrophic gastritis, and

XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be

XX used for the production of antibodies. The products can also be used for

XX detection and diagnosis.

XX SQ Sequence 288 AA;

Query Match 3.4%; Score 7; DB 19; Length 288;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 DEIKSY 190
|||||

DB 18 DEIKSY 24

RESULT 28
AAG40720
ID AAG40720 standard; Protein; 293 AA.
XX
AC AAG40720;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 50563.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130444.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132488.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149388.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 3.4%; Score 7; DB 21; Length 293;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 55 QOLDPAL 61
DB 41 QOLDPAL 47

```

```

RESULT 29
AAW88339
ID AAW88339 standard; Protein: 330 AA.
XX
XX AAW88339;
XX

```

```

DT 26-APR-1999 (first entry)
XX
DE Salmonella enterica O antigen CDP-abequose pathway protein.
XX
XX O antigen; ddhd gene; CDP-abequose; diagnosis.
XX
OS Salmonella enterica.
XX
PN MO9850531-A1.
XX
PD 12-NOV-1998.
XX
PF 01-MAY-1998; 98WC-AU00315.
XX
PR 22-JUL-1997; 97AU-0008162.
PR 01-MAY-1997; 97AU-0006545.
XX
PA (UNSY ) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX WPI, 1999-059669/05.
DR N-PSDB; AAX06751.
XX
PT Nucleic acid molecules specific for bacterial polysaccharide
PT antigens - useful for detecting specific strains in, e.g. food,
PT faeces or patient samples
XX
XX Disclosure; Fig 10; 165bp; English.
XX
XX This is the amino acid sequence of a protein of the CDP-abequose
CC pathway that is encoded by the ddhd gene of a gene cluster (see
CC AAX06751) involved in the biosynthesis of the Salmonella enterica
CC serotype B strain LT2 O antigen (see also AAW99335-50). The use of
CC nucleic acid molecules derived from assembly and transport genes,
CC particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC genes, within O antigen gene clusters improves the specificity of
CC methods for the detection and identification of O antigens, e.g. in
CC testing food- or faecal-derived samples, or samples from patients.
CC The O antigen is a major virulence factor of S. enterica.
XX
XX Sequence 330 AA;
SQ

```

```

Query Match 3.4%; Score 7; DB 20; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 57 LPAALSA 63
DB 23 LPAALSA 29

```

```

RESULT 30
AAG40719
ID AAG40719 standard; Protein: 333 AA.
XX
XX AAG40719;
AC
XX
XX 18-OCT-2000 (first entry)
DT
DE Zea mays protein fragment SEQ ID NO: 50562.
XX
XX Zea mays protein fragment SEQ ID NO: 50562.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
XX Zea mays subsp. mays.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX

```

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0122548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138099.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147320.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.4%; Score 7; DB 21; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 QQLDAL 61
|||||
DB 81 QQLDAL 87

RESULT 31

AA95047
ID AA95047 standard; Protein; 336 AA.

AC AA95047;
XX

DT 23-JUN-2000 (first entry)
XX

DE Candida albicans polypeptide sequence # 15.
XX

KM Candida albicans infection; growth; survival; medicament; AIDS;
XX

KM vulvovaginitis; immunocompromised patient; treat.
XX

OS Candida albicans.
XX

PN EP82401-A2.
XX

PD 01-MAR-2000.
XX

PF 23-DEC-1998; 98EP-0310694.
XX

PR 14-AUG-1998; 98GB-0017796.
XX

XX (JANC) JANSSEN PHARM NV.
XX

PA Contreras RH, Nelissen B, De Backer MD, Luyten WHML, Viaene JE;
PI Loghe MG;
XX

DR WPI; 2000-258614/23.
XX
XX Essential polypeptides isolated from Candida albicans, useful in the
PT treatment of diseases caused by C.albicans, especially in
PT immunocompromised subjects, e.g., AIDS patients -
XX
XX Claim 3; Page 47-48; 133pp; English.

CC This sequence represents a polypeptide that is critical for the survival
CC and growth of Candida albicans. The C. albicans nucleic acid molecules
CC encoding the polypeptides of the invention may be used as probes and
CC primers for detecting homologous nucleic acid molecule sequences. The
CC polypeptides and nucleic acid molecules and compounds identified as
CC selectively modulating the expression of the polypeptides, may be used as
CC medicaments or for the preparation of a medicament to treat C.albicans
CC associated diseases, especially in AIDS patients and to treat
CC vulvovaginitis in otherwise healthy females. The use of the polypeptides
CC and polynucleotide sequences to treat C.albicans associated diseases has
CC fewer side effects and less toxicity than previously used methods such as
CC the use of amphotericin. This method is therefore especially suitable for
CC immunocompromised patients, such as AIDS patients.

SQ Sequence 336 AA;

Query Match 3.4%; Score 7; DB 21; Length 336;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 GYEGRV 142
|||||
DB 327 GYEGRV 333

RESULT 32

AAU69429
ID AAU69429 standard; Protein; 349 AA.

AC AAU69429;
XX

DT 30-JUN-2002 (first entry)
XX

DE Lung small cell carcinoma antigen #23.
XX

KM Human; cytostatic; antitumour; lung small cell cancer antigen;
XX

KM tumour; lung cancer.
XX

OS Homo sapiens.
XX

PN WO200177168-A2.
XX

PD 18-OCT-2001.
XX

PF 11-APR-2001; 2001WO-US11859.
XX

PR 11-APR-2000; 2000US-196780P.
XX

PR 21-JUN-2000; 2000US-213361P.
XX

PR 01-SEP-2000; 2000US-229763P.
XX

PR 05-SEP-2000; 2000US-230629P.
XX

PR 14-SEP-2000; 2000US-232565P.
XX

PR 19-DEC-2000; 2000US-257037P.
XX

PR 08-JAN-2001; 2001US-260796P.
XX

XX (CORI-) CORIXA CORP.
XX

PI Lodes WJ, Wang T, Mohamath R, Indrias CV;
XX WPI; 2002-010896/01.
XX N-PSDB; AAS61669.
XX

PT Lung tumour polynucleotide and polypeptides useful in therapy and
PS diagnosis of cancer especially lung cancer -
XX Claim 2; Page 291-292; 295pp; English.

XX The invention relates to novel isolated lung small cell cancer antigen
CC polynucleotides (I) and polypeptides (II) used in a method of detecting
CC cancer in a patient. The method is optionally performed by
CC utilizing oligonucleotides (III), where the biological sample
CC from the patient is contacted with (III), detecting the amount of
CC polynucleotide hybridized to (III) in the sample and comparing the
CC amount of polynucleotide to a predetermined cut-off value and thereby
CC determining cancer in a patient (I), (II) or antigen-presenting cells
CC expressing (II) is useful for stimulating and/or expanding T cells
CC specific for a tumor protein. The method comprises contacting T cells
CC with one of the components under conditions to permit the stimulation
CC and/or expansion of the cells. A composition comprising (I) is useful for
CC stimulating an immune response in a patient and for inhibiting the
CC development of a cancer especially lung cancer in a patient. An
CC isolated T cell population is useful for removing tumor cells from the
CC biological sample and for inhibiting the development of cancer in a
CC patient. AA69407-AA69431 represent novel human lung small cell
CC cancer antigen amino acid sequences of the invention.

SQ Sequence 349 AA;

Query Match 3.4%; Score 7; DB 23; Length 349;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 KAIELNP 102
Db 152 KAIELNP 158

RESULT 33

ABG18243
ID ABG18243 standard; Protein; 350 AA.

AC ABG18243;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #18234.

KW Human; Chromosome mapping; gene mapping; gene therapy; forensic;

KW Food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

FN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS82430.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 48602; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 350 AA;

Query Match 3.4%; Score 7; DB 22; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 DAALSAD 64
Db 62 DAALSAD 68

RESULT 34

AAG40718
ID AAG40718 standard; Protein; 358 AA.

AC AAG40718;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 50561.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence; corn.

OS Zea mays subsp. mays.

FN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135533.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160981.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 3.4%; Score 7; DB 21; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 QQLDAL 61
 |||||
 Db 106 QQLDAL 112

RESULT 35

AB860104
 ID ABB60104 standard; Protein; 360 AA.

XX ABB60104;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 7104.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li FWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL04207.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 7104; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB857737-AB872072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 360 AA;

Query Match 3.4%; Score 7; DB 22; Length 360;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SACSSTP 24
 |||||
 Db 94 SACSSTP 100

RESULT 36

AA888515
 ID AAB88515 standard; Protein; 378 AA.

XX AAB88515;

XX 04-JUN-2001 (first entry)

XX Haemophilus influenzae essential bacterial protein SEQ ID NO:48.

XX Haemophilus influenzae; essential bacterial gene; identification;

XX Otitis media; meningitis; upper respiratory tract infection;

XX Infection; antimicrobial.

XX Haemophilus influenzae.

XX WO200111033-A2.

XX 15-FEB-2001.

XX 03-AUG-2000; 2000WO-US21176.

XX 04-AUG-1999; 99US-0368382.

XX (ABBO) ABBOTT LAB.

XX Chovan LE, Hessler PE, Reich KA;

XX WPI; 2001-147511/15.

XX N-PSDB; AAF94368.

XX Essential bacterial genes from Haemophilus influenzae and methods for
 PT identifying 'essential' genes that may be potential therapeutic targets
 PT -

XX Claim 9; Page 88-89; 185pp; English.

XX AAF94345 to AAF94409 represent essential bacterial genes from
 CC Haemophilus influenzae, which encode the proteins given in AAB88492 to
 CC AAB88556. The present invention also describes methods for identifying
 CC essential bacterial genes (i.e. those essential to the survival of a
 CC bacterium) using a transposition system. The methods are used to
 CC identify essential genes from bacteria, especially H. influenzae (which
 CC causes otitis media, meningitis and upper respiratory tract infections)
 CC which may be used as targets for potential antimicrobial agents.
 CC AAF94410 to AAF94416 represent PCR primers used in the exemplification
 CC of the present invention.

XX Sequence 378 AA;

Query Match 3.4%; Score 7; DB 22; Length 378;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QQLDAL 62
 |||||
 Db 221 QQLDAL 227

RESULT 37

AAU91446
 ID AAU91446 standard; Protein; 378 AA.

XX AAU91446;

XX 18-JUN-2002 (first entry)
 XX Haemophilus influenzae essential gene #24.
 DE Essential bacterial gene; antifungal agent; antibacterial agent;
 XX antiparasitic agent; insecticidal agent; microbial infection;
 KM mucous membrane infection; otitis media; sinusitis; bronchitis;
 KM alveolitis; conjunctivitis; pneumonia; meningitis; epiglottitis;
 XX cellulitis; septic arthritis.
 OS Haemophilus influenzae.
 PN W0200218601-A2.
 XX 07-MAR-2002.
 PD 22-AUG-2001; 2001MO-US26245.
 PF 25-AUG-2000; 2000US-0649145.
 XX (ABBO) ABBOTT LAB.
 PA Chovan LE, Heesler PE, Reich KA;
 PI MPI; 2002-304258/34.
 DR N-PSDB; ABRK4932.
 XX Essential bacterial genes in Haemophilus influenzae necessary for
 PT bacterium's growth and survival, useful for screening inhibitors of
 PT polypeptides and developing therapeutic agents e.g. antimicrobial
 XX
 PS Claim 9; Page 89; 185pp; English.
 CC The invention describes an essential bacterial gene (I) comprising a
 CC purified polynucleotide isolated from Haemophilus influenzae where (I)
 CC is essential to H. influenzae survival. The encoded polypeptide (II) is
 CC useful for screening substances that function to inhibit essential H.
 CC influenzae polypeptides by contacting (II) with the desired substances
 CC and measuring the response by a screen from specific, enzyme, general,
 CC affinity, phenotypic and binding screen. (I) and (II) are useful in
 CC developing therapeutic agents such as antifungal, antibacterial and
 CC antiparasitic agent, insecticidal agent, and preventive antimicrobial
 CC agents which are effective in preventing microbial infection or useful
 CC in treatment of that particular infection. (I) and (II) may also be
 CC useful in treatment of mucous membrane infections such as otitis media,
 CC sinusitis, bronchitis, alveolitis, conjunctivitis, pneumonia, meningitis,
 CC epiglottitis, cellulitis and septic arthritis. This is the amino acid
 CC sequence of an essential H. influenzae gene, described in the invention.
 XX
 SO Sequence 378 AA;
 QY Query Match 3.4%; Score 7; DB 23; Length 378;
 Db Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 56 QUDALUS 62
 Db 221 QUDALUS 227
 RESULT 38
 AAU77092
 ID AAU77092 standard; Protein; 382 AA.
 AC AAU77092;
 XX 05-JUN-2002 (first entry)
 XX Human viral protein U binding protein, Ubp.
 DE Human viral protein U binding protein, Ubp; Vpu;
 XX human immunodeficiency virus; HIV-1; CD4; Gag; anti-HIV; viral titre.
 KM

XX Homo sapiens.
 OS
 XX US2002022221-A1.
 PN 21-FEB-2002.
 PD 29-APR-1999; 99US-0301978.
 PF 30-APR-1998; 98US-083567P.
 XX
 PR (PANG/) PANGANIBAN A T.
 PA (CALL/) CALLAHAN M A.
 PA (HAND/) HANDLEY M A.
 XX
 PI Panganiban AT, Callahan MA, Handley MA;
 DR MPI; 2002-267445/31.
 DR N-PSDB; ABRK12101.
 XX
 PT New preparation of viral protein U (Vpu) binding protein (Ubp), for
 PT identifying modulators of interaction between Vpu and Ubp, and between
 PT Gag and Ubp
 XX
 PS Claim 16; Fig 1; 29pp; English.
 CC The invention relates to a preparation of viral protein U (Vpu) binding
 CC protein (Ubp), where Vpu is encoded by human immunodeficiency virus
 CC (HIV)-1 that promotes degradation of virus receptor, CD4, and enhances
 CC release of virus particles from cells. Also included are the gene
 CC sequence and protein sequence of human Ubp, a method of identifying
 CC modulators/inhibitors of the Ubp/Vpu interaction comprising: (a) exposing
 CC Vpu and Ubp together in the presence of a candidate inhibitor/modulator
 CC under conditions where Vpu and Ubp can interact when an inhibitor is not
 CC present; and (b) determining whether Vpu/Ubp interaction occurs, an
 CC inhibitor of Ubp/Vpu interaction or Gag/Ubp interaction, an anti-Ubp
 CC antibody, creating peptide fragments of the human Ubp protein and
 CC detecting members of the Ubp superfamily using probes derived from the
 CC human Ubp cDNA. Ubp is useful in an assay to identify modulators
 CC (preferably inhibitors) of the Ubp/Vpu interaction. Ubp is also useful
 CC for determining whether a candidate compound/modulates (preferably,
 CC inhibits) the interaction between Gag and Ubp. Ubp is also useful in an
 CC assay to examine the effect of Ubp on particle release. The Ubp gene is
 CC useful as an excellent candidate for evaluation of anti-HIV therapeutics.
 CC The anti-Ubp antibody is useful for detecting Ubp. Overexpression of Ubp
 CC in virus expressing cells resulted in significant reduction of HIV-1
 CC virion release, and thus Ubp plays a role in Vpu-mediated enhancement
 CC of particle release. Ubp is also useful for enhancing the release of
 CC particles to increase the titre of HIV or vectors or other retrovirus
 CC vectors. An in vitro system for detecting interaction between Vpu
 CC and Ubp, and Gag and Ubp, allows rapid and straight forward examination
 CC of mutants and reagents that affect protein interaction. The
 CC present sequence is the human Ubp protein.
 XX
 SO Sequence 382 AA;
 QY Query Match 3.4%; Score 7; DB 23; Length 382;
 Db Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 96 KALELNP 102
 Db 116 KALELNP 122
 RESULT 39
 ABG30309
 ID ABG30309 standard; Protein; 444 AA.
 AC ABG30309;
 XX 18-FEB-2002 (first entry)
 XX

DE Novel human diagnostic protein #30300.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI, 2001-639362/73.
 DR N-PSDB; AAS94496.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 60668; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 444 AA;
 XX
 QY Query Match 3.4%; Score 7; DB 22; Length 444;
 XX Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Db Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 73 TLAKYQ 79
 XX |||||
 Db 67 TLAKYQ 73
 XX
 RESULT 40
 ABG22445
 ID ABG22445 standard; Protein; 477 AA.
 XX
 AC ABG22445;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #22436.
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI, 2001-639362/73.
 DR N-PSDB; AAS86632.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 52804; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting, or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 477 AA;
 XX
 QY Query Match 3.4%; Score 7; DB 22; Length 477;
 XX Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Db Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 58 DAALSAD 64
 XX |||||
 Db 189 DAALSAD 195.
 XX
 Search completed: July 8, 2003, 11:44:32
 Job time : 93 secs


```
977 5 2.5 239 2 US-08-553-497A-18 Sequence 18, Appl
978 5 2.5 242 3 US-09-167-434-8 Sequence 8, Appl
979 5 2.5 242 3 US-08-853-755-8 Sequence 8, Appl
980 5 2.5 242 4 US-09-067-800-2 Sequence 2, Appl
981 5 2.5 242 4 US-09-105-652-2 Sequence 2, Appl
982 5 2.5 242 4 US-09-349-677-2 Sequence 2, Appl
983 5 2.5 243 1 US-08-439-725A-2 Sequence 2, Appl
984 5 2.5 243 1 US-08-439-725A-2 Sequence 8, Appl
985 5 2.5 243 1 US-08-867-471-2 Sequence 2, Appl
986 5 2.5 243 2 US-08-867-471-8 Sequence 8, Appl
987 5 2.5 243 2 US-08-438-439C-25 Sequence 25, Appl
988 5 2.5 243 2 US-08-951-822-21 Sequence 21, Appl
989 5 2.5 243 3 US-09-167-434-9 Sequence 9, Appl
990 5 2.5 243 3 US-08-705-245-1 Sequence 1, Appl
991 5 2.5 243 3 US-08-705-245-8 Sequence 8, Appl
992 5 2.5 243 3 US-08-853-755-9 Sequence 9, Appl
993 5 2.5 243 4 US-09-368-951-21 Sequence 21, Appl
994 5 2.5 243 4 US-09-469-242-2 Sequence 2, Appl
995 5 2.5 244 5 PCT-US96-06664-2 Sequence 2, Appl
996 5 2.5 244 3 US-09-135-782-4 Sequence 4, Appl
997 5 2.5 244 4 US-09-267-177-14 Sequence 14, Appl
998 5 2.5 245 2 US-08-438-439C-2 Sequence 2, Appl
999 5 2.5 245 2 US-08-951-822-24 Sequence 24, Appl
1000 5 2.5 245 3 US-08-705-245-2 Sequence 2, Appl
```

ALIGNMENTS

```
RESULT 1
US-08-841-349-4
; Sequence 4, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISRA, LOBA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO4470USO
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2154
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-841-349-4
```

```
Query Match
Best Local Similarity 3.9%; Score 8; DB 2; Length 2154;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 87 QTKAQLF 94
Db 1357 QTKAQLF 1364
```

```
RESULT 2
US-09-301-978C-15
; Sequence 15, Application US/09301978C
; Patent No. 6392015
; GENERAL INFORMATION:
; APPLICANT: Pangniban, Antonio
; APPLICANT: Callahan, Mark A.
; TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VPU and GAG
; FILE REFERENCE: 960296.95335
; CURRENT APPLICATION NUMBER: US/09/301,978C
; CURRENT FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,567
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
```

```
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus Type 1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)-(34)
; OTHER INFORMATION: TPR2
US-09-301-978C-15
```

```
Query Match
Best Local Similarity 3.4%; Score 7; DB 4; Length 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 96 KALELNP 102
Db 23 KALELNP 29
```

```
RESULT 3
US-09-301-978C-2
; Sequence 2, Application US/09301978C
; Patent No. 6392015
; GENERAL INFORMATION:
; APPLICANT: Pangniban, Antonio
; APPLICANT: Callahan, Mark A.
; TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VPU and GAG
; FILE REFERENCE: 960296.95335
; CURRENT APPLICATION NUMBER: US/09/301,978C
; CURRENT FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,567
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-301-978C-2
```

```
Query Match
Best Local Similarity 3.4%; Score 7; DB 4; Length 382;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 96 KALELNP 102
Db 116 KALELNP 122
```

```
RESULT 4
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159
```

```
Query Match
Best Local Similarity 3.4%; Score 7; DB 4; Length 10182;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 167 NNAKSAL 173
|||||
DB 3580 NNAKSAL 3586

RESULT 5
US-08-471-058-1
; Sequence 1, Application US/08471058
; Patent No. 5770443
; GENERAL INFORMATION:
; APPLICANT: Kiefer, Michael C.
; APPLICANT: Barr, Philip J.
; TITLE OF INVENTION: NOVEL APOPTOSIS MODULATING
; TITLE OF INVENTION: PROTEINS, DNA ENCODING THE PROTEINS AND METHODS OF USE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,058
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/320,157
; FILING DATE: 07-OCT-1994
; APPLICATION NUMBER: 08/160,067
; FILING DATE: 30-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lehnhardt, Susan K.
; REGISTRATION NUMBER: 33,943
; REFERENCE/DOCKET NUMBER: 23647-20007.12
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-471-058-1
Query Match 2.9%; Score 6; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2e+05; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;
QY 139 GRVAI 144
|||||
DB 3 GRVAI 8

RESULT 6
US-08-471-057-1
; Sequence 1, Application US/08471057
; Patent No. 6015687
; GENERAL INFORMATION:
; APPLICANT: KIEFER, MICHAEL C.
; APPLICANT: BARR, PHILIP J.
; TITLE OF INVENTION: NOVEL APOPTOSIS-MODULATING PROTEINS, DNA
; TITLE OF INVENTION: ENCODING THE PROTEINS AND METHODS OF USE THEREOF

NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FORSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/320,157
FILING DATE: 07-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: LEHNHARDT, SUSAN K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20007.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-471-057-1
Query Match 2.9%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2e+05; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;
QY 139 GRVAI 144
|||||
DB 3 GRVAI 8

RESULT 7
US-08-461-005-2
; Sequence 2, Application US/08461005
; Patent No. 5734035
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip J
; APPLICANT: Magland, Barry M
; TITLE OF INVENTION: Nematode Vaccine
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley and Lardner
; STREET: Suite 500, 1800 Diagonal Road
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22133-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,005
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,685
; FILING DATE: 06-OCT-1992

CLASSIFICATION: 536
APPLICATION NUMBER: AU PK4487
FILING DATE: 06-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU92/00041
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A
REGISTRATION NUMBER: 29,768
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Haemonchus contortus
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /note= "May be Leu or Val"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /note= "Amino acid unknown"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 19
OTHER INFORMATION: /note= "Amino acid unknown"
US-08-461-005-2

Query Match 2.9%; Score 6; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 AYRTLA 75
Db 8 AYRTLA 13

RESULT 8
US-07-930-685-2
Sequence 2, Application US/07930685
Patent No. 5871738
GENERAL INFORMATION:
APPLICANT: Sharp, Phillip J
APPLICANT: Magland, Barry M
TITLE OF INVENTION: Nematode Vaccine
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley and Lardner
STREET: Suite 500, 1800 Diagonal Road
CITY: Alexandria
STATE: Virginia
COUNTRY: United States of America
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/930,685
FILING DATE: 19921006
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK4487
FILING DATE: 06-FEB-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU92/00041
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A
REGISTRATION NUMBER: 29,768
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Haemonchus contortus
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /note= "May be Leu or Val"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /note= "Amino acid unknown"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 19
OTHER INFORMATION: /note= "Amino acid unknown"
US-07-930-685-2

Query Match 2.9%; Score 6; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 AYRTLA 75
Db 8 AYRTLA 13

RESULT 9
US-08-190-802A-187
Sequence 187, Application US/08190802A
Patent No. 5519003
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,802A
FILING DATE: 01-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 8600-0139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: L1S1 (human) r1, Fig. 34
US-08-190-802A-187

Query Match 2.9%; Score 6; DB 1; Length 31;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 ASEDAT 85
DB 21 ASEDAT 26

RESULT 10

US-08-477-346-187
Sequence 187, Application US/08477346
Patent No. 6262023
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: L1S1 (human) r1, Fig. 34
US-08-477-346-187

Query Match 2.9%; Score 6; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 ASEDAT 85
DB 21 ASEDAT 26

RESULT 11

US-08-473-089-187
Sequence 187, Application US/08473089
Patent No. 6342368
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,089
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: L1S1 (human) r1, Fig. 34
US-08-473-089-187

Query Match 2.9%; Score 6; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 ASEDAT 85
DB 21 ASEDAT 26

RESULT 12

US-08-487-072A-187
Sequence 187, Application US/08487072A
Patent No. 6423684
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington

STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,072A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: L1S1 (human) r1, Fig. 34

Query Match 2.9%; Score 6; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-08-487-072A-187
Db 21 ASEDAT 26
QY 80 ASEDAT 85
|||||
RESULT 13
US-09-301-978C-18
Sequence 18, Application US/09301978C
Patent No. 6392015
GENERAL INFORMATION:
APPLICANT: Pangamban, Antonio
APPLICANT: Callahan, Mark A.
TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VPU and GAG
FILE REFERENCE: 960296.95335
CURRENT APPLICATION NUMBER: US/09/301,978C
CURRENT FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 60/083,567
PRIOR FILING DATE: 1998-04-30
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 18
LENGTH: 34
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(34)
OTHER INFORMATION: TPRL
US-09-301-978C-18

Query Match 2.3%; Score 6; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 ATEINP 102
|||||
Db 24 ATEINP 29

RESULT 14
US-08-441-507-11
Sequence 11, Application US/08441507
Patent No. 6214358
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir;
APPLICANT: Smith, Penelope; and
APPLICANT: Knox, Robert Bruce
TITLE OF INVENTION: Protein Allergens of the Species Cynodon
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,507
FILING DATE: 15-May-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/969,875
FILING DATE: 30-October-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IM1-049DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 74 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-441-507-11

Query Match 2.9%; Score 6; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 1,1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 GRVAVI 144
|||||
Db 19 GRVAVI 24

RESULT 15
US-07-969-875A-11
Sequence 11, Application US/07969875A
Patent No. 6441157
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir;
APPLICANT: Smith, Penelope; and
APPLICANT: Knox, Robert Bruce
TITLE OF INVENTION: Protein Allergens of the Species
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunologic Pharmaceutical Corporation
STREET: 610 Lincoln Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA


```

;
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/969,875A
; FILING DATE: 30-October-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vanstone, Darlene A.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 041.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6010
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-969-875A-11

Query Match      2.9%; Score 6; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      139 GRVAVI 144
Db      19 GRVAVI 24

RESULT 16
US-08-441-507-13
; Sequence 13, Application US/08441507
; Patent No. 6214358
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon
; TITLE OF INVENTION: Dactylon
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LAHIVE & COCKFIELD LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,507
; FILING DATE: 15-May-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,875
; FILING DATE: 30-October-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-049DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
```

```

;
; LENGTH: 78 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-441-507-13

Query Match      2.9%; Score 6; DB 4; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      139 GRVAVI 144
Db      23 GRVAVI 28

RESULT 17
US-08-441-507-14
; Sequence 14, Application US/08441507
; Patent No. 6214358
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon
; TITLE OF INVENTION: Dactylon
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LAHIVE & COCKFIELD LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,507
; FILING DATE: 15-May-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,875
; FILING DATE: 30-October-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-049DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5-8
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15-16
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
```

LOCATION: 42
OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 71-72
OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
US-08-441-507-14

Query Match 2.9%; Score 6; DB 4; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 GRVVAI 144
DB 23 GRVVAI 28

RESULT 18
US-07-969-875A-13
Sequence 13, Application US/07969875A
Patent No. 6441157
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir;
APPLICANT: Smith, Penelope; and
APPLICANT: Knox, Robert Bruce
TITLE OF INVENTION: Protein Allergens of the Species
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunologic Pharmaceutical Corporation
STREET: 610 Lincoln Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/969,875A
FILING DATE: 30-October-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vanstone, Darlene A.,
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 041.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6010
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-969-875A-13
Query Match 2.9%; Score 6; DB 4; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 GRVVAI 144
DB 23 GRVVAI 28

RESULT 19
US-07-969-875A-14
Sequence 14, Application US/07969875A
Patent No. 6441157
GENERAL INFORMATION:

APPLICANT: Singh, Mohan Bir;
APPLICANT: Smith, Penelope; and
APPLICANT: Knox, Robert Bruce
TITLE OF INVENTION: Protein Allergens of the Species
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunologic Pharmaceutical Corporation
STREET: 610 Lincoln Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/969,875A
FILING DATE: 30-October-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vanstone, Darlene A.,
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 041.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6010
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 5-8
OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15-16
OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 42
OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 71-72
OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
US-07-969-875A-14
Query Match 2.9%; Score 6; DB 4; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 GRVVAI 144
DB 23 GRVVAI 28

RESULT 20
US-09-134-001C-3727
Sequence 3727, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al

1 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
2 FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
3 CURRENT APPLICATION NUMBER: US/09/134,001C
4 PRIOR FILING DATE: 1998-08-13
5 PRIOR APPLICATION NUMBER: US 60/064,964
6 PRIOR FILING DATE: 1997-11-08
7 PRIOR APPLICATION NUMBER: US 60/055,779
8 PRIOR FILING DATE: 1997-08-14
9 NUMBER OF SEQ ID NOS: 5674
10 SEQ ID NO 3727
11 LENGTH: 79
12 TYPE: PRT
13 ORGANISM: Staphylococcus epidermidis
14 US-09-134-001C-3727

Query Match 2.9%; Score 6; DB 4; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 LSGALI 126
DB 11 LSGALI 16

RESULT 21
US-08-353-476-75
1 Sequence 75, Application US/08353476
2 Patent No. 5871902
3 GENERAL INFORMATION:
4 APPLICANT: Weininger, Susan
5 TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A
6 TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION
7 NUMBER OF SEQUENCES: 117
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Saliwanhik & Saliwanhik
10 STREET: 2421 N.W. 41st St., Suite A-1
11 CITY: Gainesville
12 STATE: Florida
13 COUNTRY: USA
14 ZIP: 32606
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/353,476
22 FILING DATE:
23 CLASSIFICATION: 435
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Benen, Gerard H
26 REGISTRATION NUMBER: 35,746
27 REFERENCE/DOCKET NUMBER: GP-100
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (904) 375-8100
30 TELEFAX: (904) 372-5800
31 INFORMATION FOR SEQ ID NO: 75:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 85 amino acids
34 TYPE: amino acid
35 TOPOLOGY: linear
36 MOLECULE TYPE: peptide
37 HYPOTHETICAL: NO
38 ANTI-SENSE: NO
39 FRAGMENT TYPE: Internal
40 US-08-353-476-75

Query Match 2.9%; Score 6; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
DB 4 ALISGT 9

RESULT 22
US-08-679-493A-95
1 Sequence 95, Application US/08679493A
2 Patent No. 6303295
3 GENERAL INFORMATION:
4 APPLICANT: Taylor, Ethan W.
5 TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
6 FILE REFERENCE: 55-95
7 CURRENT APPLICATION NUMBER: US/08/679,493A
8 PRIOR FILING DATE: 1996-07-12
9 PRIOR APPLICATION NUMBER: 60/001203
10 PRIOR FILING DATE: 1995-07-14
11 PRIOR APPLICATION NUMBER: 60/003,112
12 PRIOR FILING DATE: 1995-09-01
13 NUMBER OF SEQ ID NOS: 216
14 SOFTWARE: Patentin Ver. 2.0
15 SEQ ID NO 95
16 LENGTH: 85
17 TYPE: PRT
18 ORGANISM: Bovine papillomavirus type 1
19 US-08-679-493A-95

Query Match 2.9%; Score 6; DB 4; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
DB 4 ALISGT 9

RESULT 23
US-08-353-476-98
1 Sequence 98, Application US/08353476
2 Patent No. 5871902
3 GENERAL INFORMATION:
4 APPLICANT: Weininger, Susan
5 TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A
6 TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION
7 NUMBER OF SEQUENCES: 117
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Saliwanhik & Saliwanhik
10 STREET: 2421 N.W. 41st St., Suite A-1
11 CITY: Gainesville
12 STATE: Florida
13 COUNTRY: USA
14 ZIP: 32606
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/353,476
22 FILING DATE:
23 CLASSIFICATION: 435
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Benen, Gerard H
26 REGISTRATION NUMBER: 35,746
27 REFERENCE/DOCKET NUMBER: GP-100
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (904) 375-8100
30 TELEFAX: (904) 372-5800
31 INFORMATION FOR SEQ ID NO: 98:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 89 amino acids
34 TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-353-476-98

Query Match 2.9%; Score 6; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
Db 8 ALISGT 13

RESULT 24
US-08-467-822-38
Sequence 38, Application US/08467822
Patent No. 5843460
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSER: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-822-38

Query Match 2.9%; Score 6; DB 2; Length 94;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 EGRVVA 143
Db 36 EGRVVA 41

RESULT 25
US-08-432-697-38
Sequence 38, Application US/08432697
Patent No. 6248330
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSER: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-432-697-38

Query Match 2.9%; Score 6; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 EGRVVA 143
Db 36 EGRVVA 41

RESULT 26
US-08-466-248-38
Sequence 38, Application US/08466248
Patent No. 6258359
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

;; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
;; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
;; TITLE OF INVENTION: POLYPEPTIDES
;; NUMBER OF SEQUENCES: 44
;; CORRESPONDENCE ADDRESSES:
;; ADDRESSES: Finnegan, Henderson, Farabow, Garrett &
;; STREET: 1300 I Street, N.W.
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/466,248
;; FILING DATE: 06-JUN-1995
;; CLASSIFICATION: 435
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/447,177
;; FILING DATE: 19-MAY-1995
;; CLASSIFICATION: 435
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/432,697
;; FILING DATE: 02-MAY-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 03495.0137-02000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 408-4000
;; TELEFAX: (202) 408-4400
;; INFORMATION FOR SEQ ID NO: 38:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 94 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-466-248-38
Query Match 2.9%; Score 6; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 EGRVVA 143
DB 36 EGRVVA 41
RESULT 27
US-09-697-367-12
; Sequence 12, Application US/09697367
; Patent No. 6333015
; GENERAL INFORMATION:
; APPLICANT: Orozco Jr., Emil M.
; APPLICANT: Caiati, Perry G.
; APPLICANT: Weng, Zude
; APPLICANT: Tarczyński, Mitchell
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: BB1166 US NA
; CURRENT APPLICATION NUMBER: US/09/697,367
; CURRENT FILING DATE: 2000-10-26
; PRIOR APPLICATION NUMBER: 60/084,529
; PRIOR FILING DATE: 1998-MAY-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-MAY-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97

;; SEQ ID NO 12
;; LENGTH: 112
;; TYPE: PRT
;; ORGANISM: Oryza sativa
US-09-697-367-12
Query Match 2.9%; Score 6; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 197 YKLSD 202
DB 15 YKLSD 20
RESULT 28
US-09-173-300-54
; Sequence 54, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173,300
; CURRENT FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 54
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Lactococcus lactis
US-09-173-300-54
Query Match 2.9%; Score 6; DB 4; Length 113;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 DYNNAK 170
DB 72 DYNNAK 77
RESULT 29
US-07-646-998A-2
; Sequence 2, Application US/07646998A
; Patent No. 5219990
; GENERAL INFORMATION:
; APPLICANT: ANDROPHY, Elliot J.
; APPLICANT: BARSDOM, James G.
; TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FISH & NEAVE
; STREET: 875 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-6250
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/646,998A
; FILING DATE: 19910128

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pierri, Margaret A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: B156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 715-0600
TELEFAX: (212) 715-0673
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-646-998A-2

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
DB 45 ALISGT 50

RESULT 30
US-07-646-998A-4
Sequence 4, Application US/07646998A
Patent No. 5219990
GENERAL INFORMATION:
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 875 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-6250
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/646,998A
FILING DATE: 19910128
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pierri, Margaret A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: B156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 715-0600
TELEFAX: (212) 715-0673
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-646-998A-4

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 176 ALISGT 181

DB 45 ALISGT 50

RESULT 31
US-07-646-998A-6
Sequence 6, Application US/07646998A
Patent No. 5219990
GENERAL INFORMATION:
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 875 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-6250
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/646,998A
FILING DATE: 19910128
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pierri, Margaret A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: B156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 715-0600
TELEFAX: (212) 715-0673
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-646-998A-6

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
DB 45 ALISGT 50

RESULT 32
US-07-646-998A-8
Sequence 8, Application US/07646998A
Patent No. 5219990
GENERAL INFORMATION:
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 875 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-6250
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/646,998A
FILING DATE: 19910128
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pierri, Margaret A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: B156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)715-0600
TELEFAX: (212)715-0673
TELEX: 14-8367
INFORMATION FOR SEQ. ID NO.: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-646-998A-8

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISCT 181
DB 45 ALISCT 50

RESULT 33
US-07-646-998A-10
Sequence 10, Application US/07646998A
Patent No. 5219990
GENERAL INFORMATION:
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
TITLE OF INVENTION: REPRESSORS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 875 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-6250
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/646,998A
FILING DATE: 19910128
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pierri, Margaret A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: B156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)715-0600
TELEFAX: (212)715-0673
TELEX: 14-8367
INFORMATION FOR SEQ. ID NO.: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein

US-07-646-998A-10

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISCT 181
DB 45 ALISCT 50

RESULT 34
US-07-646-998A-12
Sequence 12, Application US/07646998A
Patent No. 5219990
GENERAL INFORMATION:
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
TITLE OF INVENTION: REPRESSORS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 875 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-6250
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/646,998A
FILING DATE: 19910128
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pierri, Margaret A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: B156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)715-0600
TELEFAX: (212)715-0673
TELEX: 14-8367
INFORMATION FOR SEQ. ID NO.: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-646-998A-12

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISCT 181
DB 45 ALISCT 50

RESULT 35
US-07-646-998A-14
Sequence 14, Application US/07646998A
Patent No. 5219990
GENERAL INFORMATION:
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
TITLE OF INVENTION: REPRESSORS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE
STREET: 875 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-6250
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/646,998A
FILING DATE: 19910128
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pierri, Margaret A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: B156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)715-0600
TELEFAX: (212)715-0673
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-646-998A-14

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
|||||
DB 45 ALISGT 50

RESULT 36
US-08-094-128A-2
Sequence 2, Application US/08094128A
Patent No. 5595884
GENERAL INFORMATION:
APPLICANT: BARSOUM, James G.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
TITLE OF INVENTION: REPRESSORS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/094,128A
FILING DATE: 24-SEP-1993
CLASSIFICATION: 517
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/646,998
FILING DATE: 28-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00652
FILING DATE: 28-JAN-1992
ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B156CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-094-128A-2

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
|||||
DB 45 ALISGT 50

RESULT 37
US-08-094-128A-4
Sequence 4, Application US/08094128A
Patent No. 5595884
GENERAL INFORMATION:
APPLICANT: BARSOUM, James G.
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
TITLE OF INVENTION: REPRESSORS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/094,128A
FILING DATE: 24-SEP-1993
CLASSIFICATION: 517
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/646,998
FILING DATE: 28-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00652
FILING DATE: 28-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B156CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-094-128A-4

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
|||||
DB 45 ALISGT 50

RESULT 38

US-08-094-128A-6
Sequence 6, Application US/08094128A
Patent No. 5595884
GENERAL INFORMATION:
APPLICANT: BARSOUM, James G.
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
TITLE OF INVENTION: REPRESSORS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/094,128A
FILING DATE: 24-SEP-1993
CLASSIFICATION: 517
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/646,998
FILING DATE: 28-JAN-1991
PRIOR APPLICATION DATA: PCT/US92/00652
APPLICATION NUMBER: 28-JAN-1992
FILING DATE: 28-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B156CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-094-128A-6

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
|||||
DB 45 ALISGT 50

RESULT 39
US-08-094-128A-8
Sequence 8, Application US/08094128A
Patent No. 5595884
GENERAL INFORMATION:
APPLICANT: BARSOUM, James G.
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
TITLE OF INVENTION: REPRESSORS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/094,128A
FILING DATE: 24-SEP-1993
CLASSIFICATION: 517
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/646,998
FILING DATE: 28-JAN-1991
PRIOR APPLICATION DATA: PCT/US92/00652
APPLICATION NUMBER: 28-JAN-1992
FILING DATE: 28-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B156CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-094-128A-8

Query Match 2.9%; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
|||||
DB 45 ALISGT 50

RESULT 40

US-08-094-128A-10
Sequence 10, Application US/08094128A
Patent No. 5595884
GENERAL INFORMATION:
APPLICANT: BARSOUM, James G.
APPLICANT: ANDROPHY, Elliot J.
TITLE OF INVENTION: PAPILLOMAVIRUS E2 TRANS-ACTIVATION
TITLE OF INVENTION: REPRESSORS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/094,128A
FILING DATE: 24-SEP-1993
CLASSIFICATION: 517
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/646,998

; FILING DATE: 28-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00652
; FILING DATE: 28-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: B156CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-094-128A-10

Query Match 2.98; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
| | | | |
Db 45 ALISGT 50

Search completed: July 8, 2003, 11:47:48
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:47:25 ; Search time 51 Seconds
(without alignments)
465.679 Million cell updates/sec

Title: US-10-069-544-2
Perfect score: 204
Sequence: 1 MKIRVKMPMMVMMGLVLSNC.....EIKKSYDKLSDPKLSDPK 204

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 445758 seqs, 116419773 residues

Word size : 0

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Published Applications_AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US05_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	3.4	34	12 US-10-090-378-15	Sequence 15, Appl
2	7	3.4	288	9 US-09-895-913A-122	Sequence 122, Appl
3	7	3.4	349	10 US-09-833-790-434	Sequence 434, Appl
4	7	3.4	378	9 US-10-260-877-48	Sequence 48, Appl
5	7	3.4	382	12 US-10-090-378-2	Sequence 2, Appl
6	7	3.4	452	9 US-10-156-761-12413	Sequence 12413, A
7	7	3.4	547	9 US-10-156-761-12901	Sequence 12901, A
8	7	3.4	974	9 US-10-156-761-10499	Sequence 10499, A
9	6	2.9	8	9 US-10-101-482-1	Sequence 1, Appl
10	6	2.9	20	9 US-10-212-679-420	Sequence 420, Appl
11	6	2.9	20	9 US-10-212-679-421	Sequence 421, Appl
12	6	2.9	20	9 US-10-212-679-422	Sequence 422, Appl
13	6	2.9	21	9 US-10-042-431-23	Sequence 23, Appl
14	6	2.9	21	9 US-09-759-130B-393	Sequence 393, Appl
15	6	2.9	26	10 US-09-864-761-43854	Sequence 43854, A
16	6	2.9	34	9 US-10-013-939-4	Sequence 4, Appl
17	6	2.9	34	12 US-10-090-378-18	Sequence 18, Appl
18	6	2.9	45	9 US-09-733-643-18	Sequence 18, Appl
19	6	2.9	49	9 US-10-023-896-69	Sequence 69, Appl

20	6	2.9	67	9 US-10-074-095-592	Sequence 592, Appl
21	6	2.9	67	10 US-09-764-860-592	Sequence 592, Appl
22	6	2.9	68	10 US-09-864-761-43485	Sequence 43485, A
23	6	2.9	77	10 US-09-864-761-43120	Sequence 43120, A
24	6	2.9	80	9 US-09-764-891-4564	Sequence 4564, Ap
25	6	2.9	83	10 US-09-864-761-37782	Sequence 37782, A
26	6	2.9	85	7 US-08-860-844-75	Sequence 75, Appl
27	6	2.9	89	7 US-08-860-844-98	Sequence 98, Appl
28	6	2.9	112	10 US-09-918-909-12	Sequence 12, Appl
29	6	2.9	115	12 US-10-027-450-54	Sequence 54, Appl
30	6	2.9	119	9 US-10-050-704-227	Sequence 227, Appl
31	6	2.9	120	9 US-10-083-357-1281	Sequence 1281, Ap
32	6	2.9	136	9 US-10-050-704-107	Sequence 107, Appl
33	6	2.9	137	9 US-09-852-797-97	Sequence 97, Appl
34	6	2.9	137	10 US-09-853-161-97	Sequence 97, Appl
35	6	2.9	137	10 US-09-852-659A-97	Sequence 97, Appl
36	6	2.9	138	10 US-09-925-297-799	Sequence 799, Appl
37	6	2.9	152	9 US-10-156-761-11438	Sequence 11438, A
38	6	2.9	152	9 US-10-156-761-12938	Sequence 12938, A
39	6	2.9	157	9 US-09-981-286A-1	Sequence 1, Appl
40	6	2.9	162	9 US-09-852-797-63	Sequence 63, Appl
41	6	2.9	162	10 US-09-853-161-63	Sequence 63, Appl
42	6	2.9	162	10 US-09-852-659A-63	Sequence 63, Appl
43	6	2.9	165	9 US-09-981-286A-33	Sequence 33, Appl
44	6	2.9	165	9 US-09-981-286A-34	Sequence 34, Appl
45	6	2.9	169	9 US-09-981-286A-7	Sequence 7, Appl
46	6	2.9	171	9 US-10-314-058-18	Sequence 18, Appl
47	6	2.9	171	10 US-09-775-925-30	Sequence 30, Appl
48	6	2.9	171	10 US-09-847-519A-15	Sequence 15, Appl
49	6	2.9	175	9 US-10-050-704-228	Sequence 228, Appl
50	6	2.9	177	9 US-09-791-933-94	Sequence 94, Appl
51	6	2.9	178	9 US-10-091-483-198	Sequence 198, Appl
52	6	2.9	178	10 US-09-764-866-198	Sequence 198, Appl
53	6	2.9	179	9 US-09-895-913A-220	Sequence 220, Appl
54	6	2.9	182	9 US-10-156-761-11861	Sequence 11861, A
55	6	2.9	190	10 US-09-811-284-157	Sequence 157, Appl
56	6	2.9	192	9 US-09-986-468-171	Sequence 171, Appl
57	6	2.9	209	9 US-09-957-187-79	Sequence 79, Appl
58	6	2.9	209	10 US-09-815-242-13403	Sequence 13403, A
59	6	2.9	210	9 US-09-924-400-331	Sequence 331, Appl
60	6	2.9	210	9 US-10-156-761-13332	Sequence 12332, A
61	6	2.9	210	9 US-10-212-679-331	Sequence 331, Appl
62	6	2.9	210	10 US-09-810-936-331	Sequence 331, Appl
63	6	2.9	211	9 US-09-782-974C-48	Sequence 48, Appl
64	6	2.9	213	9 US-10-156-761-12947	Sequence 12947, A
65	6	2.9	216	9 US-09-738-626-5748	Sequence 5748, Ap
66	6	2.9	217	9 US-09-782-974C-4	Sequence 4, Appl
67	6	2.9	220	9 US-09-738-626-6772	Sequence 6772, Ap
68	6	2.9	226	9 US-09-965-529-26	Sequence 26, Appl
69	6	2.9	226	9 US-09-969-680A-26	Sequence 26, Appl
70	6	2.9	232	9 US-09-738-626-5583	Sequence 5583, Ap
71	6	2.9	236	10 US-09-731-872-412	Sequence 412, Appl
72	6	2.9	238	9 US-10-054-988-145	Sequence 145, Appl
73	6	2.9	238	10 US-09-739-254-145	Sequence 145, Appl
74	6	2.9	244	10 US-09-904-615-145	Sequence 145, Appl
75	6	2.9	244	10 US-09-815-242-10119	Sequence 10119, A
76	6	2.9	245	9 US-10-156-761-12301	Sequence 12301, A
77	6	2.9	251	9 US-09-895-913A-90	Sequence 90, Appl
78	6	2.9	251	9 US-10-106-696-6257	Sequence 6257, Ap
79	6	2.9	253	9 US-09-738-626-6540	Sequence 6540, Ap
80	6	2.9	262	9 US-09-738-626-4584	Sequence 4584, Ap
81	6	2.9	264	9 US-09-944-160-46	Sequence 46, Appl
82	6	2.9	264	9 US-09-944-160-47	Sequence 47, Appl
83	6	2.9	264	9 US-09-944-160-48	Sequence 48, Appl
84	6	2.9	264	9 US-09-944-160-49	Sequence 49, Appl
85	6	2.9	266	10 US-09-765-272-14	Sequence 14, Appl
86	6	2.9	267	9 US-09-764-866-736	Sequence 736, Appl
87	6	2.9	269	9 US-09-712-363-201	Sequence 201, Appl
88	6	2.9	269	10 US-09-767-041-41	Sequence 41, Appl
89	6	2.9	270	9 US-09-796-753-126	Sequence 126, Appl
90	6	2.9	270	9 US-09-905-674-2	Sequence 2, Appl
91	6	2.9	270	9 US-09-957-187-16	Sequence 16, Appl
92	6	2.9	270	9 US-09-892-877-146	Sequence 126, Appl

93	6	2.9	270	9	US-09-892-877-267	Sequence 267, App	166	6	2.9	328	9	US-10-131-823A-34	Sequence 34, Appl
94	6	2.9	270	9	US-09-948-783-137	Sequence 127, App	167	6	2.9	328	9	US-10-131-824A-34	Sequence 34, Appl
95	6	2.9	270	9	US-09-948-783-269	Sequence 269, App	168	6	2.9	328	9	US-10-131-830A-34	Sequence 34, Appl
96	6	2.9	271	9	US-10-156-761-1195	Sequence 1195, A	169	6	2.9	328	9	US-10-131-837A-34	Sequence 34, Appl
97	6	2.9	276	10	US-09-881-752A-100	Sequence 100, App	170	6	2.9	328	9	US-10-137-872A-34	Sequence 34, Appl
98	6	2.9	283	9	US-10-097-340-175	Sequence 175, App	171	6	2.9	328	9	US-10-140-860-34	Sequence 34, Appl
99	6	2.9	288	9	US-09-769-787-148	Sequence 148, App	172	6	2.9	328	9	US-10-142-417-34	Sequence 34, Appl
100	6	2.9	291	9	US-10-156-761-1163	Sequence 12163, A	173	6	2.9	328	9	US-10-147-500-34	Sequence 34, Appl
101	6	2.9	295	9	US-10-156-761-8690	Sequence 8690, App	174	6	2.9	328	9	US-10-147-502-34	Sequence 34, Appl
102	6	2.9	299	9	US-10-156-761-8363	Sequence 8363, App	175	6	2.9	328	9	US-10-147-515-34	Sequence 34, Appl
103	6	2.9	302	10	US-09-741-669-417	Sequence 417, App	176	6	2.9	328	9	US-10-147-517-34	Sequence 34, Appl
104	6	2.9	303	9	US-10-156-761-13814	Sequence 13814, A	177	6	2.9	328	9	US-10-147-519-34	Sequence 34, Appl
105	6	2.9	310	9	US-09-738-626-5018	Sequence 5018, App	178	6	2.9	328	9	US-10-147-526-34	Sequence 34, Appl
106	6	2.9	310	9	US-09-738-626-3669	Sequence 3669, App	179	6	2.9	328	9	US-10-147-527-34	Sequence 34, Appl
107	6	2.9	315	9	US-09-738-626-7000	Sequence 7000, App	180	6	2.9	328	9	US-10-152-395-34	Sequence 34, Appl
108	6	2.9	315	9	US-10-156-761-12810	Sequence 12810, A	181	6	2.9	328	9	US-10-157-782-34	Sequence 34, Appl
109	6	2.9	316	10	US-09-815-242-5152	Sequence 5152, App	182	6	2.9	328	9	US-10-121-040-34	Sequence 34, Appl
110	6	2.9	320	10	US-09-815-242-11721	Sequence 11721, A	183	6	2.9	328	9	US-10-121-056-34	Sequence 34, Appl
111	6	2.9	328	9	US-10-040-416-4	Sequence 4, Appl	184	6	2.9	328	9	US-10-121-061-34	Sequence 34, Appl
112	6	2.9	328	9	US-10-040-416-6	Sequence 6, Appl	185	6	2.9	328	9	US-10-123-235-34	Sequence 34, Appl
113	6	2.9	328	9	US-10-028-072-34	Sequence 34, Appl	186	6	2.9	328	9	US-10-124-811-34	Sequence 34, Appl
114	6	2.9	328	9	US-10-071-338-16	Sequence 16, Appl	187	6	2.9	328	9	US-10-125-926A-34	Sequence 34, Appl
115	6	2.9	328	9	US-10-121-049-34	Sequence 34, Appl	188	6	2.9	328	9	US-10-125-930A-34	Sequence 34, Appl
116	6	2.9	328	9	US-10-123-904-34	Sequence 34, Appl	189	6	2.9	328	9	US-10-127-831A-34	Sequence 34, Appl
117	6	2.9	328	9	US-10-140-470-34	Sequence 34, Appl	190	6	2.9	328	9	US-10-127-831A-34	Sequence 34, Appl
118	6	2.9	328	9	US-10-175-746-34	Sequence 34, Appl	191	6	2.9	328	9	US-10-127-837A-34	Sequence 34, Appl
119	6	2.9	328	9	US-10-175-918-34	Sequence 34, Appl	192	6	2.9	328	9	US-10-127-838B-34	Sequence 34, Appl
120	6	2.9	328	9	US-10-176-921-34	Sequence 34, Appl	193	6	2.9	328	9	US-10-127-842A-34	Sequence 34, Appl
121	6	2.9	328	9	US-10-176-921-34	Sequence 34, Appl	194	6	2.9	328	9	US-10-127-843A-34	Sequence 34, Appl
122	6	2.9	328	9	US-10-137-865-34	Sequence 34, Appl	195	6	2.9	328	9	US-10-127-845A-34	Sequence 34, Appl
123	6	2.9	328	9	US-10-140-474-34	Sequence 34, Appl	196	6	2.9	328	9	US-10-127-846A-34	Sequence 34, Appl
124	6	2.9	328	9	US-10-142-431-34	Sequence 34, Appl	197	6	2.9	328	9	US-10-127-848A-34	Sequence 34, Appl
125	6	2.9	328	9	US-10-143-114-34	Sequence 34, Appl	198	6	2.9	328	9	US-10-127-849A-34	Sequence 34, Appl
126	6	2.9	328	9	US-10-140-002-34	Sequence 34, Appl	199	6	2.9	328	9	US-10-127-850A-34	Sequence 34, Appl
127	6	2.9	328	9	US-10-142-419-34	Sequence 34, Appl	200	6	2.9	328	9	US-10-127-851A-34	Sequence 34, Appl
128	6	2.9	328	9	US-10-123-262-34	Sequence 34, Appl	201	6	2.9	328	9	US-10-128-684A-34	Sequence 34, Appl
129	6	2.9	328	9	US-10-142-423-34	Sequence 34, Appl	202	6	2.9	328	9	US-10-128-686A-34	Sequence 34, Appl
130	6	2.9	328	9	US-10-121-050-34	Sequence 34, Appl	203	6	2.9	328	9	US-10-128-690A-34	Sequence 34, Appl
131	6	2.9	328	9	US-10-141-755-34	Sequence 34, Appl	204	6	2.9	328	9	US-10-128-691A-34	Sequence 34, Appl
132	6	2.9	328	9	US-10-143-032-34	Sequence 34, Appl	205	6	2.9	328	9	US-10-128-691A-34	Sequence 34, Appl
133	6	2.9	328	9	US-10-160-501-27	Sequence 27, Appl	206	6	2.9	328	9	US-10-131-828A-34	Sequence 34, Appl
134	6	2.9	328	9	US-10-123-108-34	Sequence 34, Appl	207	6	2.9	328	9	US-10-131-836A-34	Sequence 34, Appl
135	6	2.9	328	9	US-10-123-261-34	Sequence 34, Appl	208	6	2.9	328	9	US-10-137-868-34	Sequence 34, Appl
136	6	2.9	328	9	US-10-140-921-34	Sequence 34, Appl	209	6	2.9	328	9	US-10-146-729-34	Sequence 34, Appl
137	6	2.9	328	9	US-10-140-928-34	Sequence 34, Appl	210	6	2.9	328	9	US-10-146-791-34	Sequence 34, Appl
138	6	2.9	328	9	US-10-121-045-34	Sequence 34, Appl	211	6	2.9	328	9	US-10-147-484-34	Sequence 34, Appl
139	6	2.9	328	9	US-10-123-292-34	Sequence 34, Appl	212	6	2.9	328	9	US-10-147-492-34	Sequence 34, Appl
140	6	2.9	328	9	US-10-123-903-34	Sequence 34, Appl	213	6	2.9	328	9	US-10-147-508-34	Sequence 34, Appl
141	6	2.9	328	9	US-10-124-819-34	Sequence 34, Appl	214	6	2.9	328	9	US-10-147-512-34	Sequence 34, Appl
142	6	2.9	328	9	US-10-124-822-34	Sequence 34, Appl	215	6	2.9	328	9	US-10-158-782-34	Sequence 34, Appl
143	6	2.9	328	9	US-10-140-925-34	Sequence 34, Appl	216	6	2.9	328	9	US-10-175-735-34	Sequence 34, Appl
144	6	2.9	328	9	US-10-160-498-34	Sequence 34, Appl	217	6	2.9	328	9	US-10-175-905-34	Sequence 34, Appl
145	6	2.9	328	9	US-10-121-041-34	Sequence 34, Appl	218	6	2.9	328	9	US-10-123-907-34	Sequence 34, Appl
146	6	2.9	328	9	US-10-121-043-34	Sequence 34, Appl	219	6	2.9	328	9	US-10-124-815-34	Sequence 34, Appl
147	6	2.9	328	9	US-10-121-047-34	Sequence 34, Appl	220	6	2.9	328	9	US-10-125-921A-34	Sequence 34, Appl
148	6	2.9	328	9	US-10-123-215-34	Sequence 34, Appl	221	6	2.9	328	9	US-10-127-821A-34	Sequence 34, Appl
149	6	2.9	328	9	US-10-123-902-34	Sequence 34, Appl	222	6	2.9	328	9	US-10-127-822A-34	Sequence 34, Appl
150	6	2.9	328	9	US-10-123-908-34	Sequence 34, Appl	223	6	2.9	328	9	US-10-127-822A-34	Sequence 34, Appl
151	6	2.9	328	9	US-10-123-909-34	Sequence 34, Appl	224	6	2.9	328	9	US-10-127-824A-34	Sequence 34, Appl
152	6	2.9	328	9	US-10-123-910-34	Sequence 34, Appl	225	6	2.9	328	9	US-10-127-826A-34	Sequence 34, Appl
153	6	2.9	328	9	US-10-124-813-34	Sequence 34, Appl	226	6	2.9	328	9	US-10-127-827A-34	Sequence 34, Appl
154	6	2.9	328	9	US-10-124-817-34	Sequence 34, Appl	227	6	2.9	328	9	US-10-127-828A-34	Sequence 34, Appl
155	6	2.9	328	9	US-10-124-824-34	Sequence 34, Appl	228	6	2.9	328	9	US-10-127-832A-34	Sequence 34, Appl
156	6	2.9	328	9	US-10-125-922-34	Sequence 34, Appl	229	6	2.9	328	9	US-10-127-832A-34	Sequence 34, Appl
157	6	2.9	328	9	US-10-125-924-34	Sequence 34, Appl	230	6	2.9	328	9	US-10-127-833A-34	Sequence 34, Appl
158	6	2.9	328	9	US-10-127-825A-34	Sequence 34, Appl	231	6	2.9	328	9	US-10-127-834A-34	Sequence 34, Appl
159	6	2.9	328	9	US-10-127-829A-34	Sequence 34, Appl	232	6	2.9	328	9	US-10-127-836A-34	Sequence 34, Appl
160	6	2.9	328	9	US-10-127-835A-34	Sequence 34, Appl	233	6	2.9	328	9	US-10-127-841A-34	Sequence 34, Appl
161	6	2.9	328	9	US-10-127-839A-34	Sequence 34, Appl	234	6	2.9	328	9	US-10-127-844A-34	Sequence 34, Appl
162	6	2.9	328	9	US-10-127-901A-34	Sequence 34, Appl	235	6	2.9	328	9	US-10-128-687A-34	Sequence 34, Appl
163	6	2.9	328	9	US-10-128-693A-34	Sequence 34, Appl	236	6	2.9	328	9	US-10-128-688A-34	Sequence 34, Appl
164	6	2.9	328	9	US-10-131-813A-34	Sequence 34, Appl	237	6	2.9	328	9	US-10-128-699A-34	Sequence 34, Appl
165	6	2.9	328	9	US-10-131-818A-34	Sequence 34, Appl	238	6	2.9	328	9	US-10-128-694A-34	Sequence 34, Appl
												US-10-131-825A-34	Sequence 34, Appl

239	6	2.9	328	9	US-10-230-417-34	Sequence 34, Appl	312	6	2.9	363	9	US-10-176-985-590	Sequence 590, App
240	6	2.9	328	9	US-10-121-051-34	Sequence 34, Appl	313	6	2.9	363	9	US-10-176-987-590	Sequence 590, App
241	6	2.9	328	9	US-10-131-815A-34	Sequence 34, Appl	314	6	2.9	363	9	US-10-176-991-590	Sequence 590, App
242	6	2.9	328	9	US-10-131-817A-34	Sequence 34, Appl	315	6	2.9	363	9	US-10-176-992-590	Sequence 590, App
243	6	2.9	328	9	US-10-131-821A-34	Sequence 34, Appl	316	6	2.9	363	9	US-10-176-993-590	Sequence 590, App
244	6	2.9	328	9	US-10-131-822A-34	Sequence 34, Appl	317	6	2.9	363	9	US-10-184-658-590	Sequence 590, App
245	6	2.9	328	9	US-10-131-828A-34	Sequence 34, Appl	318	6	2.9	363	9	US-10-173-697-590	Sequence 590, App
246	6	2.9	328	9	US-10-131-835A-34	Sequence 34, Appl	319	6	2.9	363	9	US-10-173-699-590	Sequence 590, App
247	6	2.9	328	9	US-10-137-864A-34	Sequence 34, Appl	320	6	2.9	363	9	US-10-173-705-590	Sequence 590, App
248	6	2.9	328	9	US-10-137-869A-34	Sequence 34, Appl	321	6	2.9	363	9	US-10-174-576-590	Sequence 590, App
249	6	2.9	328	9	US-10-147-523-34	Sequence 34, Appl	322	6	2.9	363	9	US-10-174-585-590	Sequence 590, App
250	6	2.9	328	9	US-10-158-785-34	Sequence 34, Appl	323	6	2.9	363	9	US-10-174-586-590	Sequence 590, App
251	6	2.9	328	9	US-10-121-042-34	Sequence 34, Appl	324	6	2.9	363	9	US-10-175-747-590	Sequence 590, App
252	6	2.9	328	9	US-10-123-912-34	Sequence 34, Appl	325	6	2.9	363	9	US-10-176-481-590	Sequence 590, App
253	6	2.9	328	9	US-10-192-007-34	Sequence 34, Appl	326	6	2.9	363	9	US-10-176-485-590	Sequence 590, App
254	6	2.9	328	9	US-10-194-359-34	Sequence 34, Appl	327	6	2.9	363	9	US-10-176-487-590	Sequence 590, App
255	6	2.9	328	9	US-10-288-985-16	Sequence 16, Appl	328	6	2.9	363	9	US-10-176-493-590	Sequence 590, App
256	6	2.9	328	9	US-10-069-434-3	Sequence 3, Appl	329	6	2.9	363	9	US-10-176-756-590	Sequence 590, App
257	6	2.9	328	9	US-10-127-847A-34	Sequence 34, Appl	330	6	2.9	363	9	US-10-176-911-590	Sequence 590, App
258	6	2.9	328	9	US-10-156-761-9643	Sequence 643, Ap	331	6	2.9	363	9	US-10-176-919-590	Sequence 590, App
259	6	2.9	328	9	US-10-175-590-34	Sequence 34, Appl	332	6	2.9	363	9	US-10-176-925-590	Sequence 590, App
260	6	2.9	328	9	US-10-289-980-16	Sequence 16, Appl	333	6	2.9	363	9	US-10-176-978-590	Sequence 590, App
261	6	2.9	330	9	US-10-040-416-2	Sequence 2, Appl	334	6	2.9	363	9	US-10-179-510-590	Sequence 590, App
262	6	2.9	330	10	US-09-800-487A-2	Sequence 2, Appl	335	6	2.9	363	9	US-10-180-543-590	Sequence 590, App
263	6	2.9	333	9	US-09-595-225-2	Sequence 2, Appl	336	6	2.9	363	9	US-10-180-544-590	Sequence 590, App
264	6	2.9	333	9	US-09-791-932-91	Sequence 91, Appl	337	6	2.9	363	9	US-10-180-546-590	Sequence 590, App
265	6	2.9	333	9	US-09-791-932-114	Sequence 114, Appl	338	6	2.9	363	9	US-10-180-547-590	Sequence 590, App
266	6	2.9	333	9	US-09-990-940-4	Sequence 4, Appl	339	6	2.9	363	9	US-10-180-548-590	Sequence 590, App
267	6	2.9	333	9	US-10-225-567A-680	Sequence 680, App	340	6	2.9	363	9	US-10-180-551-590	Sequence 590, App
268	6	2.9	333	10	US-09-895-686-3	Sequence 3, Appl	341	6	2.9	363	9	US-10-180-552-590	Sequence 590, App
269	6	2.9	334	10	US-09-815-242-11070	Sequence 11070, A	342	6	2.9	363	9	US-10-181-000-590	Sequence 590, App
270	6	2.9	336	9	US-09-940-244-165	Sequence 165, App	343	6	2.9	363	9	US-10-183-010-590	Sequence 590, App
271	6	2.9	336	9	US-09-864-636A-550	Sequence 550, App	344	6	2.9	363	9	US-10-183-012-590	Sequence 590, App
272	6	2.9	336	9	US-09-864-636A-553	Sequence 553, App	345	6	2.9	363	9	US-10-184-614-590	Sequence 590, App
273	6	2.9	336	10	US-09-815-242-11899	Sequence 11899, A	346	6	2.9	363	9	US-10-184-621-590	Sequence 590, App
274	6	2.9	336	10	US-09-777-430A-50	Sequence 30, Appl	347	6	2.9	363	9	US-10-184-623-590	Sequence 590, App
275	6	2.9	337	9	US-09-791-932-115	Sequence 115, Appl	348	6	2.9	363	9	US-10-184-635-590	Sequence 590, App
276	6	2.9	337	9	US-10-225-567A-591	Sequence 591, Appl	349	6	2.9	363	9	US-10-184-637-590	Sequence 590, App
277	6	2.9	338	9	US-09-738-626-4405	Sequence 4405, Ap	350	6	2.9	363	9	US-10-184-644-590	Sequence 590, App
278	6	2.9	342	10	US-09-991-258-10	Sequence 10, Appl	351	6	2.9	363	9	US-10-184-652-590	Sequence 590, App
279	6	2.9	346	10	US-09-815-242-11437	Sequence 11437, A	352	6	2.9	363	9	US-10-187-594-590	Sequence 590, App
280	6	2.9	346	10	US-09-815-242-11656	Sequence 11656, A	353	6	2.9	363	9	US-10-187-596-590	Sequence 590, App
281	6	2.9	351	9	US-09-820-0968-2	Sequence 2, Appl	354	6	2.9	363	9	US-10-187-598-590	Sequence 590, App
282	6	2.9	351	9	US-09-940-244-341	Sequence 341, Appl	355	6	2.9	363	9	US-10-187-745-590	Sequence 590, App
283	6	2.9	351	10	US-09-820-296-2	Sequence 2, Appl	356	6	2.9	363	9	US-10-187-886-590	Sequence 590, App
284	6	2.9	352	9	US-10-079-384-20	Sequence 20, Appl	357	6	2.9	363	9	US-10-189-464-590	Sequence 590, App
285	6	2.9	354	9	US-09-095-478-6	Sequence 6, Appl	358	6	2.9	363	9	US-10-176-751-590	Sequence 590, App
286	6	2.9	361	10	US-09-796-858-28	Sequence 28, Appl	359	6	2.9	363	9	US-10-176-760-590	Sequence 590, App
287	6	2.9	363	9	US-10-080-960-8	Sequence 8, Appl	360	6	2.9	363	9	US-10-176-990-590	Sequence 590, App
288	6	2.9	363	9	US-10-174-590-590	Sequence 590, App	361	6	2.9	363	9	US-10-180-541-590	Sequence 590, App
289	6	2.9	363	9	US-10-176-758-590	Sequence 590, App	362	6	2.9	363	9	US-10-180-546-590	Sequence 590, App
290	6	2.9	363	9	US-10-175-737-590	Sequence 590, App	363	6	2.9	363	9	US-10-180-548-590	Sequence 590, App
291	6	2.9	363	9	US-10-166-221-2	Sequence 2, Appl	364	6	2.9	363	9	US-10-180-551-590	Sequence 590, App
292	6	2.9	363	9	US-10-173-706-590	Sequence 590, App	365	6	2.9	363	9	US-10-180-998-590	Sequence 590, App
293	6	2.9	363	9	US-10-175-738-590	Sequence 590, App	366	6	2.9	363	9	US-10-180-999-590	Sequence 590, App
294	6	2.9	363	9	US-10-175-752-590	Sequence 590, App	367	6	2.9	363	9	US-10-183-013-590	Sequence 590, App
295	6	2.9	363	9	US-10-176-482-590	Sequence 590, App	368	6	2.9	363	9	US-10-184-612-590	Sequence 590, App
296	6	2.9	363	9	US-10-176-757-590	Sequence 590, App	369	6	2.9	363	9	US-10-184-614-590	Sequence 590, App
297	6	2.9	363	9	US-10-176-913-590	Sequence 590, App	370	6	2.9	363	9	US-10-184-617-590	Sequence 590, App
298	6	2.9	363	9	US-10-180-552-590	Sequence 590, App	371	6	2.9	363	9	US-10-184-622-590	Sequence 590, App
299	6	2.9	363	9	US-10-180-557-590	Sequence 590, App	372	6	2.9	363	9	US-10-184-628-590	Sequence 590, App
300	6	2.9	363	9	US-10-173-700-590	Sequence 590, App	373	6	2.9	363	9	US-10-184-629-590	Sequence 590, App
301	6	2.9	363	9	US-10-174-572-590	Sequence 590, App	374	6	2.9	363	9	US-10-184-632-590	Sequence 590, App
302	6	2.9	363	9	US-10-174-579-590	Sequence 590, App	375	6	2.9	363	9	US-10-184-631-590	Sequence 590, App
303	6	2.9	363	9	US-10-174-582-590	Sequence 590, App	376	6	2.9	363	9	US-10-184-633-590	Sequence 590, App
304	6	2.9	363	9	US-10-174-588-590	Sequence 590, App	377	6	2.9	363	9	US-10-184-636-590	Sequence 590, App
305	6	2.9	363	9	US-10-175-739-590	Sequence 590, App	378	6	2.9	363	9	US-10-184-640-590	Sequence 590, App
306	6	2.9	363	9	US-10-175-740-590	Sequence 590, App	379	6	2.9	363	9	US-10-184-650-590	Sequence 590, App
307	6	2.9	363	9	US-10-175-743-590	Sequence 590, App	380	6	2.9	363	9	US-10-184-651-590	Sequence 590, App
308	6	2.9	363	9	US-10-176-488-590	Sequence 590, App	381	6	2.9	363	9	US-10-187-588-590	Sequence 590, App
309	6	2.9	363	9	US-10-176-492-590	Sequence 590, App	382	6	2.9	363	9	US-10-187-597-590	Sequence 590, App
310	6	2.9	363	9	US-10-176-747-590	Sequence 590, App	383	6	2.9	363	9	US-10-187-598-590	Sequence 590, App
311	6	2.9	363	9	US-10-176-750-590	Sequence 590, App	384	6	2.9	363	9	US-10-187-600-590	Sequence 590, App

385	6	2.9	363	9	US-10-187-601-590	Sequence 590, App	458	6	2.9	363	9	US-10-187-750-590	Sequence 590, App
386	6	2.9	363	9	US-10-187-602-590	Sequence 590, App	459	6	2.9	363	9	US-10-188-780-590	Sequence 590, App
387	6	2.9	363	9	US-10-187-603-590	Sequence 590, App	460	6	2.9	363	9	US-10-192-015-590	Sequence 590, App
388	6	2.9	363	9	US-10-187-741-590	Sequence 590, App	461	6	2.9	363	9	US-10-194-394-590	Sequence 590, App
389	6	2.9	363	9	US-10-187-743-590	Sequence 590, App	462	6	2.9	363	9	US-10-194-482-590	Sequence 590, App
390	6	2.9	363	9	US-10-187-746-590	Sequence 590, App	463	6	2.9	363	9	US-10-194-485-590	Sequence 590, App
391	6	2.9	363	9	US-10-187-747-590	Sequence 590, App	464	6	2.9	363	9	US-10-195-885-590	Sequence 590, App
392	6	2.9	363	9	US-10-187-751-590	Sequence 590, App	465	6	2.9	363	9	US-10-195-889-590	Sequence 590, App
393	6	2.9	363	9	US-10-187-753-590	Sequence 590, App	466	6	2.9	363	9	US-10-195-899-590	Sequence 590, App
394	6	2.9	363	9	US-10-187-754-590	Sequence 590, App	467	6	2.9	363	9	US-10-196-748-590	Sequence 590, App
395	6	2.9	363	9	US-10-187-757-590	Sequence 590, App	468	6	2.9	363	9	US-10-196-750-590	Sequence 590, App
396	6	2.9	363	9	US-10-187-884-590	Sequence 590, App	469	6	2.9	363	9	US-10-197-699-590	Sequence 590, App
397	6	2.9	363	9	US-10-188-767-590	Sequence 590, App	470	6	2.9	363	9	US-10-197-700-590	Sequence 590, App
398	6	2.9	363	9	US-10-188-769-590	Sequence 590, App	471	6	2.9	363	9	US-10-197-705-590	Sequence 590, App
399	6	2.9	363	9	US-10-188-770-590	Sequence 590, App	472	6	2.9	363	9	US-10-197-708-590	Sequence 590, App
400	6	2.9	363	9	US-10-188-773-590	Sequence 590, App	473	6	2.9	363	9	US-10-198-764-590	Sequence 590, App
401	6	2.9	363	9	US-10-188-781-590	Sequence 590, App	474	6	2.9	363	9	US-10-198-766-590	Sequence 590, App
402	6	2.9	363	9	US-10-194-361-590	Sequence 590, App	475	6	2.9	363	9	US-10-198-765-590	Sequence 590, App
403	6	2.9	363	9	US-10-194-423-590	Sequence 590, App	476	6	2.9	363	9	US-10-198-768-590	Sequence 590, App
404	6	2.9	363	9	US-10-195-897-590	Sequence 590, App	477	6	2.9	363	9	US-10-198-769-590	Sequence 590, App
405	6	2.9	363	9	US-10-195-901-590	Sequence 590, App	478	6	2.9	363	9	US-10-199-305-590	Sequence 590, App
406	6	2.9	363	9	US-10-196-756-590	Sequence 590, App	479	6	2.9	363	9	US-10-199-306-590	Sequence 590, App
407	6	2.9	363	9	US-10-173-708-590	Sequence 590, App	480	6	2.9	363	9	US-10-199-310-590	Sequence 590, App
408	6	2.9	363	9	US-10-176-479-590	Sequence 590, App	481	6	2.9	363	9	US-10-199-311-590	Sequence 590, App
409	6	2.9	363	9	US-10-176-748-590	Sequence 590, App	482	6	2.9	363	9	US-10-199-314-590	Sequence 590, App
410	6	2.9	363	9	US-10-176-916-590	Sequence 590, App	483	6	2.9	363	9	US-10-199-317-590	Sequence 590, App
411	6	2.9	363	9	US-10-179-507-590	Sequence 590, App	484	6	2.9	363	9	US-10-199-666-590	Sequence 590, App
412	6	2.9	363	9	US-10-179-516-590	Sequence 590, App	485	6	2.9	363	9	US-10-199-666-590	Sequence 590, App
413	6	2.9	363	9	US-10-179-519-590	Sequence 590, App	486	6	2.9	363	9	US-10-199-666-590	Sequence 590, App
414	6	2.9	363	9	US-10-179-525-590	Sequence 590, App	487	6	2.9	363	9	US-10-201-534-590	Sequence 590, App
415	6	2.9	363	9	US-10-180-540-590	Sequence 590, App	488	6	2.9	363	9	US-10-201-770-590	Sequence 590, App
416	6	2.9	363	9	US-10-180-545-590	Sequence 590, App	489	6	2.9	363	9	US-10-201-855-590	Sequence 590, App
417	6	2.9	363	9	US-10-183-006-590	Sequence 590, App	490	6	2.9	363	9	US-10-201-855-590	Sequence 590, App
418	6	2.9	363	9	US-10-183-008-590	Sequence 590, App	491	6	2.9	363	9	US-10-202-467-590	Sequence 590, App
419	6	2.9	363	9	US-10-183-017-590	Sequence 590, App	492	6	2.9	363	9	US-10-202-476-590	Sequence 590, App
420	6	2.9	363	9	US-10-183-019-590	Sequence 590, App	493	6	2.9	363	9	US-10-202-934-590	Sequence 590, App
421	6	2.9	363	9	US-10-184-618-590	Sequence 590, App	494	6	2.9	363	9	US-10-202-935-590	Sequence 590, App
422	6	2.9	363	9	US-10-184-625-590	Sequence 590, App	495	6	2.9	363	9	US-10-202-936-590	Sequence 590, App
423	6	2.9	363	9	US-10-184-626-590	Sequence 590, App	496	6	2.9	363	9	US-10-202-939-590	Sequence 590, App
424	6	2.9	363	9	US-10-184-627-590	Sequence 590, App	497	6	2.9	363	9	US-10-205-504-590	Sequence 590, App
425	6	2.9	363	9	US-10-184-645-590	Sequence 590, App	498	6	2.9	363	9	US-10-205-508-590	Sequence 590, App
426	6	2.9	363	9	US-10-184-654-590	Sequence 590, App	499	6	2.9	363	9	US-10-205-895-590	Sequence 590, App
427	6	2.9	363	9	US-10-184-655-590	Sequence 590, App	500	6	2.9	363	9	US-10-205-899-590	Sequence 590, App
428	6	2.9	363	9	US-10-184-655-590	Sequence 590, App	501	6	2.9	363	9	US-10-205-900-590	Sequence 590, App
429	6	2.9	363	9	US-10-188-774-590	Sequence 590, App	502	6	2.9	363	9	US-10-205-909-590	Sequence 590, App
430	6	2.9	363	9	US-10-188-775-590	Sequence 590, App	503	6	2.9	363	9	US-10-183-002-590	Sequence 590, App
431	6	2.9	363	9	US-10-194-462-590	Sequence 590, App	504	6	2.9	363	9	US-10-184-621-590	Sequence 590, App
432	6	2.9	363	9	US-10-195-900-590	Sequence 590, App	505	6	2.9	363	9	US-10-184-638-590	Sequence 590, App
433	6	2.9	363	9	US-10-196-743-590	Sequence 590, App	506	6	2.9	363	9	US-10-187-752-590	Sequence 590, App
434	6	2.9	363	9	US-10-196-745-590	Sequence 590, App	507	6	2.9	363	9	US-10-187-755-590	Sequence 590, App
435	6	2.9	363	9	US-10-196-760-590	Sequence 590, App	508	6	2.9	363	9	US-10-187-887-590	Sequence 590, App
436	6	2.9	363	9	US-10-196-762-590	Sequence 590, App	509	6	2.9	363	9	US-10-194-461-590	Sequence 590, App
437	6	2.9	363	9	US-10-197-695-590	Sequence 590, App	510	6	2.9	363	9	US-10-195-892-590	Sequence 590, App
438	6	2.9	363	9	US-10-176-484-590	Sequence 590, App	511	6	2.9	363	9	US-10-196-751-590	Sequence 590, App
439	6	2.9	363	9	US-10-176-753-590	Sequence 590, App	512	6	2.9	363	9	US-10-197-694-590	Sequence 590, App
440	6	2.9	363	9	US-10-176-917-590	Sequence 590, App	513	6	2.9	363	9	US-10-197-707-590	Sequence 590, App
441	6	2.9	363	9	US-10-176-982-590	Sequence 590, App	514	6	2.9	363	9	US-10-197-709-590	Sequence 590, App
442	6	2.9	363	9	US-10-179-506-590	Sequence 590, App	515	6	2.9	363	9	US-10-199-303-590	Sequence 590, App
443	6	2.9	363	9	US-10-179-513-590	Sequence 590, App	516	6	2.9	363	9	US-10-199-318-590	Sequence 590, App
444	6	2.9	363	9	US-10-179-514-590	Sequence 590, App	517	6	2.9	363	9	US-10-199-458-590	Sequence 590, App
445	6	2.9	363	9	US-10-179-522-590	Sequence 590, App	518	6	2.9	363	9	US-10-199-462-590	Sequence 590, App
446	6	2.9	363	9	US-10-180-556-590	Sequence 590, App	519	6	2.9	363	9	US-10-201-324-590	Sequence 590, App
447	6	2.9	363	9	US-10-180-560-590	Sequence 590, App	520	6	2.9	363	9	US-10-201-328-590	Sequence 590, App
448	6	2.9	363	9	US-10-183-015-590	Sequence 590, App	521	6	2.9	363	9	US-10-201-537-590	Sequence 590, App
449	6	2.9	363	9	US-10-184-615-590	Sequence 590, App	522	6	2.9	363	9	US-10-201-538-590	Sequence 590, App
450	6	2.9	363	9	US-10-184-620-590	Sequence 590, App	523	6	2.9	363	9	US-10-201-549-590	Sequence 590, App
451	6	2.9	363	9	US-10-184-643-590	Sequence 590, App	524	6	2.9	363	9	US-10-201-530-590	Sequence 590, App
452	6	2.9	363	9	US-10-184-656-590	Sequence 590, App	525	6	2.9	363	9	US-10-202-408-590	Sequence 590, App
453	6	2.9	363	9	US-10-192-010-590	Sequence 590, App	526	6	2.9	363	9	US-10-202-409-590	Sequence 590, App
454	6	2.9	363	9	US-10-195-894-590	Sequence 590, App	527	6	2.9	363	9	US-10-202-411-590	Sequence 590, App
455	6	2.9	363	9	US-10-205-908-590	Sequence 590, App	528	6	2.9	363	9	US-10-202-472-590	Sequence 590, App
456	6	2.9	363	9	US-10-184-619-590	Sequence 590, App	529	6	2.9	363	9	US-10-205-502-590	Sequence 590, App
457	6	2.9	363	9	US-10-186-855-590	Sequence 590, App	530	6	2.9	363	9	US-10-205-507-590	Sequence 590, App
					US-10-187-599-590	Sequence 590, App						US-10-205-511-590	Sequence 590, App

677	6	2.9	363	9	US-10-179-526-590	Sequence 590, App	750	6	2.9	393	10	US-09-925-297-825	Sequence 825, App
678	6	2.9	363	9	US-10-173-701-590	Sequence 590, App	751	6	2.9	394	9	US-09-924-400-336	Sequence 336, App
679	6	2.9	363	9	US-10-179-511-590	Sequence 590, App	752	6	2.9	394	9	US-10-212-679-336	Sequence 336, App
680	6	2.9	363	9	US-10-179-518-590	Sequence 590, App	753	6	2.9	396	9	US-10-156-761-7760	Sequence 7760, App
681	6	2.9	363	9	US-10-183-018-590	Sequence 590, App	754	6	2.9	398	7	US-08-809-423A-56	Sequence 56, App
682	6	2.9	363	9	US-10-184-624-590	Sequence 590, App	755	6	2.9	399	10	US-09-893-737-10	Sequence 10, App
683	6	2.9	363	9	US-10-184-657-590	Sequence 590, App	756	6	2.9	401	9	US-10-156-761-7673	Sequence 7673, App
684	6	2.9	363	9	US-10-197-701-590	Sequence 590, App	757	6	2.9	407	10	US-09-815-244-11965	Sequence 11965, A
685	6	2.9	363	9	US-10-197-706-590	Sequence 590, App	758	6	2.9	410	10	US-09-815-244-11993	Sequence 11993, A
686	6	2.9	363	9	US-10-201-857-590	Sequence 590, App	759	6	2.9	410	10	US-09-815-244-13628	Sequence 13628, A
687	6	2.9	363	9	US-10-202-413-590	Sequence 590, App	760	6	2.9	410	10	US-09-847-081B-4	Sequence 4, App
688	6	2.9	363	9	US-10-202-938-590	Sequence 590, App	761	6	2.9	411	9	US-10-101-464A-982	Sequence 982, App
689	6	2.9	363	9	US-10-202-940-590	Sequence 590, App	762	6	2.9	411	9	US-10-036-959B-8	Sequence 8, App
690	6	2.9	363	9	US-10-205-508-590	Sequence 590, App	763	6	2.9	412	9	US-09-095-478-3	Sequence 3, App
691	6	2.9	363	9	US-10-205-905-590	Sequence 590, App	764	6	2.9	413	9	US-10-156-761-12114	Sequence 12414, A
692	6	2.9	363	9	US-10-206-918-590	Sequence 590, App	765	6	2.9	416	9	US-10-156-761-10859	Sequence 10859, A
693	6	2.9	363	9	US-10-208-025-590	Sequence 590, App	766	6	2.9	420	9	US-10-156-761-9543	Sequence 9543, App
694	6	2.9	363	9	US-10-198-760-590	Sequence 590, App	767	6	2.9	423	9	US-09-917-384-4	Sequence 4, App
695	6	2.9	363	9	US-10-201-772-590	Sequence 590, App	768	6	2.9	423	9	US-09-917-383-9	Sequence 9, App
696	6	2.9	363	9	US-10-225-567A-555	Sequence 555, App	769	6	2.9	423	9	US-09-917-383-4	Sequence 4, App
697	6	2.9	363	9	US-10-184-613-590	Sequence 590, App	770	6	2.9	423	9	US-09-917-383-9	Sequence 9, App
698	6	2.9	363	9	US-10-187-729-590	Sequence 590, App	771	6	2.9	426	9	US-10-013-802A-8	Sequence 8, App
699	6	2.9	363	9	US-10-206-907-590	Sequence 590, App	772	6	2.9	426	9	US-09-738-626-4624	Sequence 4624, App
700	6	2.9	363	9	US-10-183-009-590	Sequence 590, App	773	6	2.9	426	9	US-09-095-478-1	Sequence 1, App
701	6	2.9	363	9	US-10-187-755-590	Sequence 590, App	774	6	2.9	429	9	US-10-103-313-566	Sequence 566, App
702	6	2.9	363	10	US-09-736-131-2	Sequence 2, App	775	6	2.9	432	9	US-10-156-761-10067	Sequence 10067, A
703	6	2.9	365	10	US-09-908-711-100	Sequence 100, App	776	6	2.9	432	9	US-09-738-626-4678	Sequence 4678, App
704	6	2.9	365	10	US-09-908-711-100	Sequence 100, App	777	6	2.9	432	9	US-09-738-626-6081	Sequence 6081, App
705	6	2.9	368	9	US-09-770-509-24	Sequence 24, App	778	6	2.9	439	9	US-10-288-222A-18	Sequence 18, App
706	6	2.9	372	9	US-09-984-245-171	Sequence 171, App	779	6	2.9	440	9	US-09-866-570A-45	Sequence 45, App
707	6	2.9	372	9	US-09-966-262-171	Sequence 171, App	780	6	2.9	440	9	US-10-063-547-134	Sequence 134, App
708	6	2.9	372	9	US-09-963-966-171	Sequence 171, App	781	6	2.9	440	9	US-10-174-590-386	Sequence 386, App
709	6	2.9	372	9	US-10-143-090-171	Sequence 171, App	782	6	2.9	440	9	US-10-174-590-386	Sequence 386, App
710	6	2.9	373	7	US-08-911-824-52	Sequence 52, App	783	6	2.9	440	9	US-10-176-758-386	Sequence 386, App
711	6	2.9	374	9	US-09-728-626-5901	Sequence 6501, App	784	6	2.9	440	9	US-10-063-616-134	Sequence 134, App
712	6	2.9	374	9	US-10-300-819-4	Sequence 4, App	785	6	2.9	440	9	US-10-175-737-386	Sequence 386, App
713	6	2.9	374	9	US-10-156-761-13423	Sequence 13423, A	786	6	2.9	440	9	US-10-063-502-134	Sequence 134, App
714	6	2.9	375	10	US-09-815-242-11803	Sequence 11803, A	787	6	2.9	440	9	US-10-173-706-386	Sequence 386, App
715	6	2.9	379	9	US-09-095-478-8	Sequence 8, App	788	6	2.9	440	9	US-10-175-753-386	Sequence 386, App
716	6	2.9	380	9	US-10-128-714-3056	Sequence 3056, App	789	6	2.9	440	9	US-10-176-482-386	Sequence 386, App
717	6	2.9	382	9	US-09-920-923-5	Sequence 5, App	790	6	2.9	440	9	US-10-176-757-386	Sequence 386, App
718	6	2.9	382	10	US-09-993-844-5	Sequence 5, App	791	6	2.9	440	9	US-10-176-913-386	Sequence 386, App
719	6	2.9	382	10	US-09-947-267-7	Sequence 7, App	792	6	2.9	440	9	US-10-180-552-386	Sequence 386, App
720	6	2.9	384	9	US-09-924-400-304	Sequence 304, App	793	6	2.9	440	9	US-10-180-552-386	Sequence 386, App
721	6	2.9	384	9	US-09-924-400-326	Sequence 326, App	794	6	2.9	440	9	US-10-173-700-386	Sequence 386, App
722	6	2.9	384	9	US-09-924-400-332	Sequence 332, App	795	6	2.9	440	9	US-10-173-700-386	Sequence 386, App
723	6	2.9	384	9	US-09-924-400-333	Sequence 333, App	796	6	2.9	440	9	US-10-174-572-386	Sequence 386, App
724	6	2.9	384	9	US-09-924-400-340	Sequence 340, App	797	6	2.9	440	9	US-10-174-582-386	Sequence 386, App
725	6	2.9	384	9	US-09-924-400-340	Sequence 340, App	798	6	2.9	440	9	US-10-174-582-386	Sequence 386, App
726	6	2.9	384	9	US-09-443-745-17	Sequence 17, App	799	6	2.9	440	9	US-10-175-739-386	Sequence 386, App
727	6	2.9	384	9	US-10-292-525-12	Sequence 12, App	800	6	2.9	440	9	US-10-175-740-386	Sequence 386, App
728	6	2.9	384	9	US-10-235-567A-130	Sequence 130, App	801	6	2.9	440	9	US-10-175-743-386	Sequence 386, App
729	6	2.9	384	9	US-10-212-679-304	Sequence 304, App	802	6	2.9	440	9	US-10-176-488-386	Sequence 386, App
730	6	2.9	384	9	US-10-212-679-326	Sequence 326, App	803	6	2.9	440	9	US-10-176-492-386	Sequence 386, App
731	6	2.9	384	9	US-10-212-679-332	Sequence 332, App	804	6	2.9	440	9	US-10-176-747-386	Sequence 386, App
732	6	2.9	384	9	US-10-212-679-333	Sequence 333, App	805	6	2.9	440	9	US-10-176-750-386	Sequence 386, App
733	6	2.9	384	9	US-10-212-679-334	Sequence 334, App	806	6	2.9	440	9	US-10-176-985-386	Sequence 386, App
734	6	2.9	384	9	US-10-212-679-340	Sequence 340, App	807	6	2.9	440	9	US-10-176-985-386	Sequence 386, App
735	6	2.9	384	10	US-09-160-116-18	Sequence 18, App	808	6	2.9	440	9	US-10-176-981-386	Sequence 386, App
736	6	2.9	384	10	US-09-825-301-8	Sequence 8, App	809	6	2.9	440	9	US-10-176-992-386	Sequence 386, App
737	6	2.9	384	10	US-09-810-936-304	Sequence 304, App	810	6	2.9	440	9	US-10-176-992-386	Sequence 386, App
738	6	2.9	384	10	US-09-810-936-326	Sequence 326, App	811	6	2.9	440	9	US-10-176-993-386	Sequence 386, App
739	6	2.9	384	10	US-09-810-936-332	Sequence 332, App	812	6	2.9	440	9	US-10-184-658-386	Sequence 386, App
740	6	2.9	384	10	US-09-810-936-332	Sequence 332, App	813	6	2.9	440	9	US-10-184-658-386	Sequence 386, App
741	6	2.9	384	10	US-09-810-936-333	Sequence 333, App	814	6	2.9	440	9	US-10-177-697-386	Sequence 386, App
742	6	2.9	384	10	US-09-810-936-334	Sequence 334, App	815	6	2.9	440	9	US-10-177-697-386	Sequence 386, App
743	6	2.9	384	10	US-09-429-755-304	Sequence 304, App	816	6	2.9	440	9	US-10-174-585-386	Sequence 386, App
744	6	2.9	390	9	US-09-971-536-53	Sequence 53, App	817	6	2.9	440	9	US-10-174-585-386	Sequence 386, App
745	6	2.9	390	9	US-09-966-782A-18	Sequence 18, App	818	6	2.9	440	9	US-10-174-585-386	Sequence 386, App
746	6	2.9	390	9	US-09-443-745-18	Sequence 18, App	819	6	2.9	440	9	US-10-175-747-386	Sequence 386, App
747	6	2.9	390	9	US-10-292-525-13	Sequence 13, App	820	6	2.9	440	9	US-10-175-747-386	Sequence 386, App
748	6	2.9	390	9	US-10-225-567A-199	Sequence 199, App	821	6	2.9	440	9	US-10-176-481-386	Sequence 386, App
749	6	2.9	391	9	US-09-160-116-19	Sequence 19, App	822	6	2.9	440	9	US-10-176-481-386	Sequence 386, App
					US-10-156-761-10550	Sequence 10550, A							


```
969 6 2.9 440 9 US-10-198-764-386 Sequence 386, App
970 6 2.9 440 9 US-10-198-765-386 Sequence 386, App
971 6 2.9 440 9 US-10-198-768-386 Sequence 386, App
972 6 2.9 440 9 US-10-198-769-386 Sequence 386, App
973 6 2.9 440 9 US-10-199-305-386 Sequence 386, App
974 6 2.9 440 9 US-10-199-306-386 Sequence 386, App
975 6 2.9 440 9 US-10-199-310-386 Sequence 386, App
976 6 2.9 440 9 US-10-199-311-386 Sequence 386, App
977 6 2.9 440 9 US-10-199-314-386 Sequence 386, App
978 6 2.9 440 9 US-10-199-317-386 Sequence 386, App
979 6 2.9 440 9 US-10-199-665-386 Sequence 386, App
980 6 2.9 440 9 US-10-199-666-386 Sequence 386, App
981 6 2.9 440 9 US-10-199-669-386 Sequence 386, App
982 6 2.9 440 9 US-10-201-534-386 Sequence 386, App
983 6 2.9 440 9 US-10-201-770-386 Sequence 386, App
984 6 2.9 440 9 US-10-201-855-386 Sequence 386, App
985 6 2.9 440 9 US-10-201-856-386 Sequence 386, App
986 6 2.9 440 9 US-10-202-469-386 Sequence 386, App
987 6 2.9 440 9 US-10-202-470-386 Sequence 386, App
988 6 2.9 440 9 US-10-202-476-386 Sequence 386, App
989 6 2.9 440 9 US-10-202-934-386 Sequence 386, App
990 6 2.9 440 9 US-10-202-935-386 Sequence 386, App
991 6 2.9 440 9 US-10-202-936-386 Sequence 386, App
992 6 2.9 440 9 US-10-202-939-386 Sequence 386, App
993 6 2.9 440 9 US-10-205-504-386 Sequence 386, App
994 6 2.9 440 9 US-10-205-509-386 Sequence 386, App
995 6 2.9 440 9 US-10-205-895-386 Sequence 386, App
996 6 2.9 440 9 US-10-205-899-386 Sequence 386, App
997 6 2.9 440 9 US-10-205-900-386 Sequence 386, App
998 6 2.9 440 9 US-10-205-909-386 Sequence 386, App
999 6 2.9 440 9 US-10-227-693-134 Sequence 134, App
1000 6 2.9 440 9 US-10-183-002-386 Sequence 386, App
```

ALIGNMENTS

```
RESULT 1
US-10-090-378-15
; Sequence 15, Application US/10090378
; Patent No. US20020115830A1
; GENERAL INFORMATION:
; APPLICANT: Pangnabhan, Antonito
; APPLICANT: Callahan, Mark A.
; TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VP1 and GAG
; TITLE OF INVENTION: Interaction with U Binding Protein (UBP)
; FILE REFERENCE: 960296.95335
; CURRENT APPLICATION NUMBER: US/10/090.378
; CURRENT FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: US/09/301.978C
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,567
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus Type 1
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(34)
; OTHER INFORMATION: TPR2
US-10-090-378-15
```

```
Query Match 3.4%; Score 7; DB 12; Length 34;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 96 KAIEINP 102
| | | | |
Db 23 KAIEINP 29
```

```
RESULT 2
US-09-895-913A-122
; Sequence 122, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Omen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895.913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 122
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-122
```

```
Query Match 3.4%; Score 7; DB 9; Length 288;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 184 DEIKSY 190
| | | | |
Db 18 DEIKSY 24
```

```
RESULT 3
US-09-833-790-434
; Sequence 434, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongrong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indrias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833.790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 434
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-790-434
```

```
Query Match 3.4%; Score 7; DB 10; Length 349;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 96 KAIEINP 102
| | | | |
Db 152 KAIEINP 158
```

```
RESULT 4
US-10-260-877-48
; Sequence 48, Application US/10260877
; Publication No. US20030021813A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Hessler, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; FILE REFERENCE: 6565.US.P1
; CURRENT APPLICATION NUMBER: US/10/260,877
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/649,145
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 48
; LENGTH: 378
; TYPE: PRT
; ORGANISM: H. Influenzae
US-10-260-877-48

```

```

Query Match      3.4%; Score 7; DB 9; Length 378;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      56 OLDPAALS 62
DB      221 OLDPAALS 227

```

```

RESULT 5
US-10-090-378-2
; Sequence 2, Application US/10090378
; Patent No. US20020115830A1
; GENERAL INFORMATION:
; APPLICANT: Pangniban, Antonio
; APPLICANT: Callahan, Mark A.
; TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VP1 and GAG
; TITLE OF INVENTION: Interaction with U Binding Protein (UBP)
; FILE REFERENCE: 960296.95335
; CURRENT APPLICATION NUMBER: US/10/090,378
; CURRENT FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: US/09/301,978C
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,567
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-378-2

```

```

Query Match      3.4%; Score 7; DB 12; Length 382;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      96 KALELNP 102
DB      116 KALELNP 122

```

```

RESULT 6
US-10-156-761-12413
; Sequence 12413, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI

```

```

; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 12413
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12413

```

```

Query Match      3.4%; Score 7; DB 9; Length 452;
Best Local Similarity 100.0%; Pred.No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      57 LDALSA 63
DB      334 LDALSA 340

```

```

RESULT 7
US-10-156-761-12901
; Sequence 12901, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 12901
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12901

```

```

Query Match      3.4%; Score 7; DB 9; Length 547;
Best Local Similarity 100.0%; Pred.No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      133 RAIGYEG 139
DB      226 RAIGYEG 232

```

```

RESULT 8
US-10-156-761-10499
; Sequence 10499, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI

```

APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10499
LENGTH: 974
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-10499

Query Match 3.4%; Score 7; DB 9; Length 974;
Best Local Similarity 100.0%; Pred. No. 2,6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 LDALSA.63
|||||
Db 314 LDALSA 320

RESULT 9
US-10-101-482-1
Sequence 1, Application US/10101482
Publication No. US2003008837A1
GENERAL INFORMATION:
APPLICANT: KIEFER, MICHAEL C.
BARR, PHILIP J.

TITLE OF INVENTION: NOVEL APOPTOSIS-MODULATING PROTEIN, DNA
ENCODING THE PROTEINS AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/101,482
FILING DATE: 18-Mar-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/320,157
FILING DATE: 07-OCT-1994

ATTORNEY/AGENT INFORMATION:
NAME: LEHNHARDT, SUSAN K.

REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20007, 20

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 8-amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-101-482-1

Query Match 2.9%; Score 6; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 4e+05;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 GRVVAI 144
|||||
Db 3 GRVVAI 8

RESULT 10
US-10-212-679-420
Sequence 420, Application US/10212679
Publication No. US20030125536A1
GENERAL INFORMATION:
APPLICANT: Fanger, Gary
APPLICANT: Hirst, Shannon Kathleen
APPLICANT: Dillon, Davin
APPLICANT: Foy, Teresa
APPLICANT: Houghton, Ray
APPLICANT: Persing, David
APPLICANT: Kalos, Michael

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.419C14
CURRENT APPLICATION NUMBER: US/10/212,679
CURRENT FILING DATE: 2002-08-02
NUMBER OF SEQ ID NOS: 428

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 420
LENGTH: 20
TYPE: PRT

ORGANISM: Homo sapiens
US-10-212-679-420

Query Match 2.9%; Score 6; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 LLSDYK 198
|||||
Db 15 LLSDYK 20

RESULT 11
US-10-212-679-421

Sequence 421, Application US/10212679
Publication No. US20030125536A1
GENERAL INFORMATION:
APPLICANT: Fanger, Gary
APPLICANT: Hirst, Shannon Kathleen
APPLICANT: Dillon, Davin
APPLICANT: Foy, Teresa
APPLICANT: Houghton, Ray
APPLICANT: Persing, David
APPLICANT: Kalos, Michael

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.419C14
CURRENT APPLICATION NUMBER: US/10/212,679
CURRENT FILING DATE: 2002-08-02
NUMBER OF SEQ ID NOS: 428

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 421
LENGTH: 20
TYPE: PRT

ORGANISM: Homo sapiens
US-10-212-679-421

Query Match 2.9%; Score 6; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 LLSDYK 198
|||||
Db 10 LLSDYK 15

RESULT 12
US-10-212-679-422
; Sequence 422, Application US/10212679
; Publication No. US20030125536A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary
; APPLICANT: Hirt, Shannon Kathleen
; APPLICANT: Dillon, Davin
; APPLICANT: Foy, Teresa
; APPLICANT: Houghton, Ray
; APPLICANT: Petersing, David
; APPLICANT: Kalos, Michael
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C14
; CURRENT APPLICATION NUMBER: US/10/212,679
; CURRENT FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 422
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-212-679-422

Query Match 2.9%; Score 6; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 LLSYK 198
Db 5 LLSYK 10

RESULT 13
US-10-042-431-23
; Sequence 23, Application US/10042431
; Publication No. US20020182675A1
; GENERAL INFORMATION:
; APPLICANT: MCCARTHY, Sean A
; APPLICANT: BARNES, Thomas M
; APPLICANT: FRASER, Christopher C
; APPLICANT: SHARP, John D
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
; TITLE OF INVENTION: PREVENTIVE, THERAPEUTIC, AND OTHER USES
; FILE REFERENCE: 10147-6U2
; CURRENT APPLICATION NUMBER: US/10/042,431
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-431-23

Query Match 2.9%; Score 6; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 LSGALI 126
Db 12 LSGALI 17

RESULT 14
US-09-759-130B-393

; Sequence 393, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Frazer, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirt, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350NMIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 393
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-393

Query Match 2.9%; Score 6; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 LSGALI 126
Db 12 LSGALI 17

RESULT 15
US-09-864-761-43854
; Sequence 43854, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312

```

; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43854
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002090.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: O70354, EVALUATE 2.00e-03
; OTHER INFORMATION: EST_HUMAN HIT: H10914.1, EVALUATE 6.00e-08
; US-09-864-761-43854

Query Match      2.9%; Score 6; DB 10; Length 26;
Best Local Similarity 100.0%; Pred.No. 97;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      156 EAAKSP 161
Db      16 EAAKSP 21

RESULT 16
; Sequence 4, Application US/10013939
; Patent No. US20020177212A1
; GENERAL INFORMATION:
; APPLICANT: Patterson, Winston C.
; APPLICANT: Ballinger, Carol A.
; TITLE OF INVENTION: Polypeptide That Interacts With Heat Shock Proteins
; FILE REFERENCE: 265.00010130
; CURRENT APPLICATION NUMBER: US/10/013,939
; CURRENT FILING DATE: 2002-06-07
```

```

; PRIOR APPLICATION NUMBER: 60/134,433
; PRIOR FILING DATE: 1999-05-17
; PRIOR APPLICATION NUMBER: 09/573,473
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-013-939-4

Query Match      2.9%; Score 6; DB 9; Length 34;
Best Local Similarity 100.0%; Pred.No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      97 AIELNP 102
Db      27 AIELNP 32

RESULT 17
; US-10-090-378-18
; Sequence 18, Application US/10090378
; Patent No. US20020115830A1
; GENERAL INFORMATION:
; APPLICANT: Pangniban, Antonio
; APPLICANT: Callahan, Mark A.
; TITLE OF INVENTION: Method of Identifying Modulators of HIV-1 VPU and GAG
; FILE REFERENCE: 960296.95335
; CURRENT APPLICATION NUMBER: US/10/090,378
; CURRENT FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: US/09/301,978C
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,567
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(34)
; OTHER INFORMATION: TPPI1
; US-10-090-378-18

Query Match      2.9%; Score 6; DB 12; Length 34;
Best Local Similarity 100.0%; Pred.No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      97 AIELNP 102
Db      24 AIELNP 29

RESULT 18
; US-09-733-643-18
; Sequence 18, Application US/09733643
; Publication No. US20030115627A1
; GENERAL INFORMATION:
; APPLICANT: Laroche, Andre J.
; APPLICANT: Huang, Timothy Y.
; APPLICANT: Frick, Michele M.
; APPLICANT: Lu, Zhen-Xiang
; APPLICANT: Huang, Hung Chang
; APPLICANT: Cheng, Kuo Joan
; TITLE OF INVENTION: Coniotyrium militans beta-(1,3) exoglucanase gene
; FILE REFERENCE: 24014US1
; CURRENT APPLICATION NUMBER: US/09/733,643
```

;; CURRENT FILING DATE: 2000-12-08
;; PRIOR APPLICATION NUMBER: US 60/170,168
;; PRIOR FILING DATE: 1999-12-10
;; NUMBER OF SEQ ID NOS: 21
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 18
;; LENGTH: 45
;; TYPE: PRT
;; ORGANISM: Bacteriophage phi-29
;; FEATURE:
;; OTHER INFORMATION: Phi-29
US-09-733-643-18

Query Match 2.9%; Score 6; DB 9; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 FEKAT 99
Db 19 FEKAT 24

RESULT 19
US-10-023-896-69
;; Sequence 69, Application US/10023896
;; Publication No. US2003002776A1
;; GENERAL INFORMATION:
;; APPLICANT: Victor Roebke
;; TITLE OF INVENTION: 29 Human Cancer Associated Proteins
;; FILE REFERENCE: PA004P1
;; CURRENT APPLICATION NUMBER: US/10/023,896
;; PRIOR FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: unassigned
;; PRIOR FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: PCT/US00/23794
;; PRIOR FILING DATE: 2000-08-30
;; PRIOR APPLICATION NUMBER: 60/152,296
;; PRIOR FILING DATE: 1999-09-03
;; PRIOR APPLICATION NUMBER: 60/158,003
;; PRIOR FILING DATE: 1999-10-06
;; NUMBER OF SEQ ID NOS: 138
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 69
;; LENGTH: 49
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-023-896-69

Query Match 2.9%; Score 6; DB 9; Length 49;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 GSTPI 26
Db 11 GSTPI 16

RESULT 20
US-10-074-095-592
;; Sequence 592, Application US/10074095
;; Publication No. US2003007704A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC008C1
;; CURRENT APPLICATION NUMBER: US/10/074,095
;; CURRENT FILING DATE: 2002-02-14
;; PRIOR APPLICATION NUMBER: 09/764,860
;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 60/179,065
;; PRIOR FILING DATE: 2000-01-31
;; PRIOR APPLICATION NUMBER: 60/180,628
;; PRIOR FILING DATE: 2000-02-04

;; PRIOR APPLICATION NUMBER: 60/214,886
;; PRIOR FILING DATE: 2000-06-28
;; PRIOR APPLICATION NUMBER: 60/217,487
;; PRIOR FILING DATE: 2000-07-11
;; PRIOR APPLICATION NUMBER: 60/225,758
;; PRIOR FILING DATE: 2000-08-14
;; PRIOR APPLICATION NUMBER: 60/220,963
;; PRIOR FILING DATE: 2000-07-26
;; PRIOR APPLICATION NUMBER: 60/217,496
;; PRIOR FILING DATE: 2000-07-11
;; PRIOR APPLICATION NUMBER: 60/225,447
;; PRIOR FILING DATE: 2000-08-14
;; PRIOR APPLICATION NUMBER: 60/218,290
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: 60/225,757
;; PRIOR FILING DATE: 2000-08-14
;; PRIOR APPLICATION NUMBER: 60/226,868
;; PRIOR FILING DATE: 2000-08-22
;; PRIOR APPLICATION NUMBER: 60/216,647
;; PRIOR FILING DATE: 2000-07-07
;; PRIOR APPLICATION NUMBER: 60/225,267
;; PRIOR FILING DATE: 2000-08-14
;; PRIOR APPLICATION NUMBER: 60/216,880
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: 60/234,274
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: 60/234,223
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: 60/228,924
;; PRIOR FILING DATE: 2000-08-30
;; PRIOR APPLICATION NUMBER: 60/224,518
;; PRIOR FILING DATE: 2000-08-14
;; PRIOR APPLICATION NUMBER: 60/236,369
;; PRIOR FILING DATE: 2000-09-29
;; PRIOR APPLICATION NUMBER: 60/224,519
;; PRIOR FILING DATE: 2000-08-14
;; PRIOR APPLICATION NUMBER: 60/220,964
;; PRIOR FILING DATE: 2000-07-26
;; PRIOR APPLICATION NUMBER: 60/241,809
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/249,299
;; PRIOR FILING DATE: 2000-11-17
;; PRIOR APPLICATION NUMBER: 60/236,327
;; PRIOR FILING DATE: 2000-09-29
;; PRIOR APPLICATION NUMBER: 60/241,785
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/244,617
;; PRIOR FILING DATE: 2000-11-01
;; PRIOR APPLICATION NUMBER: 60/225,268
;; PRIOR FILING DATE: 2000-08-14
;; PRIOR APPLICATION NUMBER: 60/236,368
;; PRIOR FILING DATE: 2000-09-29
;; PRIOR APPLICATION NUMBER: 60/251,856
;; PRIOR FILING DATE: 2000-12-08
;; PRIOR APPLICATION NUMBER: 60/251,868
;; PRIOR FILING DATE: 2000-12-08
;; PRIOR APPLICATION NUMBER: 60/229,344
;; PRIOR FILING DATE: 2000-09-01
;; PRIOR APPLICATION NUMBER: 60/234,997
;; PRIOR FILING DATE: 2000-09-25
;; PRIOR APPLICATION NUMBER: 60/229,343
;; PRIOR FILING DATE: 2000-09-01
;; PRIOR APPLICATION NUMBER: 60/229,345
;; PRIOR FILING DATE: 2000-09-01
;; PRIOR APPLICATION NUMBER: 60/229,287
;; PRIOR FILING DATE: 2000-09-01
;; PRIOR APPLICATION NUMBER: 60/229,513

```

PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/240,960
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,759
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,214
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/230,438
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/215,135
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 60/225,266
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/249,218
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,208
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,213
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,212
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,207
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,245
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,244
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,217
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215
PRIOR FILING DATE: 2000-11-17

PRIOR APPLICATION NUMBER: 60/249,264
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/232,400
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,414
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,244
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,064
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/233,063
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,397
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,399
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/241,808
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,826
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,786
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,221
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,475
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/231,243
PRIOR FILING DATE: 2000-09-08

Query Match      2.9%; Score 6; DB 9; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      89 KAORLP 94
DB      56 KAORLP 61

RESULT 21
US-09-764-860-592
Sequence 592, Application US/09764860
Patent No. US20020094953A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC008
CURRENT APPLICATION NUMBER: US/09/764,860
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1198
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 592
LENGTH: 67
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (32)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (34)
```



```
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43120
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ACO11402.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
; OTHER INFORMATION: SWISSPROT HIT: P79235, EVALUE 3.00e-05
; OTHER INFORMATION: EST_HUMAN HIT: AA203362.1, EVALUE 4.00e-05
US-09-864-761-43120

Query Match
Best Local Similarity 2.9%; Score 6; DB 10; Length 77;
Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ALERL 177
DB 65 ALERL 70

RESULT 24
US-09-764-891-4564
; Sequence 4564, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4564
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-4564

Query Match
Best Local Similarity 2.9%; Score 6; DB 9; Length 80;
Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 LAQRT 38
DB 42 LAQRT 47

RESULT 25
US-09-864-761-37782
; Sequence 37782, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
```

```

; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37782
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL021579.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HB100, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EST_HUMAN HIT: AW37534.1, EVALUE 6.00e-26
; OTHER INFORMATION: SWISSPROT HIT: P48634, EVALUE 2.00e-06
US-09-864-761-37782

Query Match
Best Local Similarity 2.9%; Score 6; DB 10; Length 83;
Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 MOSYMD 110
DB 61 MOSYMD 66

RESULT 26
US-08-860-844-75
; Sequence 75, Application US/08860844
; Publication No. US20030104361A1
; GENERAL INFORMATION:
; APPLICANT: Weininger, Susan
; APPLICANT: Weininger, Arthur M
; TITLE OF INVENTION: METHOD OF DETECTION OF NUCLEIC ACIDS WITH A
; TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st St., Suite A-1
; CITY: Gainesville
```

STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,844
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/353,476
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: GP-100C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-860-844-75

Query Match
Best Local Similarity 100.0%; Score 6; DB 7; Length 85;
Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
|||||
Db 4 ALISGT 9

RESULT 27
US-08-860-844-98
Sequence 98, Application US/08860844
Publication No. US20030104361A1
GENERAL INFORMATION:
APPLICANT: Weininger, Susan
APPLICANT: Weininger, Arthur M.
TITLE OF INVENTION: METHOD OF DETECTION OF NUCLEIC ACIDS WITH A
TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION
NUMBER OF SEQUENCES: 118
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st St., Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,844
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/353,476
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: GP-100C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 98:
SEQUENCE CHARACTERISTICS:
LENGTH: 89 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-860-844-98

Query Match
Best Local Similarity 100.0%; Score 6; DB 7; Length 89;
Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ALISGT 181
|||||
Db 8 ALISGT 13

RESULT 28
US-09-918-909-12
Sequence 12, Application US/09918909
Patent No. US20020090704A1
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Cairni, Perry G.
APPLICANT: Orozco Jr., Emil M.
TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
FILE REFERENCE: B81166 US CIP
CURRENT APPLICATION NUMBER: US/09/918,909
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/084,529
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: PCT/US99/09865
PRIOR FILING DATE: 1999-05-06
PRIOR APPLICATION NUMBER: 09/697,367
PRIOR FILING DATE: 2000-10-26
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Microsoft Office 97
SEQ ID NO 12
LENGTH: 112
TYPE: PRT
ORGANISM: Oryza sativa
US-09-918-909-12

Query Match
Best Local Similarity 100.0%; Score 6; DB 10; Length 112;
Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 YKLSD 202
|||||
Db 15 YKLSD 20

RESULT 29
US-10-027-450-54
Sequence 54, Application US/10027450
Patent No. US20020102715A1
GENERAL INFORMATION:
APPLICANT: Falco, Saverio Carl
APPLICANT: Hitz, William D.
APPLICANT: Kinney, Anthony J.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Rafalski, J. Antoni
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES

FILE REFERENCE: BB-1126
CURRENT APPLICATION NUMBER: US/10/027,450
CURRENT FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/063,423
PRIOR FILING DATE: 1997 October 28
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 54
LENGTH: 113
TYPE: PRT
ORGANISM: Lactococcus lactis
US-10-027-450-54

Query Match 2.9%; Score 6; DB 12; Length 115;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 DYNNAK 170
DB 72 DYNNAK 77

RESULT 30
US-10-050-704-227
Sequence 227, Application US/10050704
Publication No. US20030050442A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 62 Human Secreted Proteins
FILE REFERENCE: P2039P1
CURRENT APPLICATION NUMBER: US/10/050,704
CURRENT FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: 09/684,524
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: PCT/US00/08979
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/128,693
PRIOR FILING DATE: 1999-04-09
PRIOR APPLICATION NUMBER: 60/130,991
PRIOR FILING DATE: 1999-04-26
NUMBER OF SEQ ID NOS: 344
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 227
LENGTH: 119
TYPE: PRT
ORGANISM: Homo sapiens
US-10-050-704-227

Query Match 2.9%; Score 6; DB 9; Length 119;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 SALERA 176
DB 41 SALERA 46

RESULT 31
US-10-083-357-1281
Sequence 1281, Application US/10083357
Publication No. US20030054370A1
GENERAL INFORMATION:
APPLICANT: Qiangdong Zeng et al.
TITLE OF INVENTION: Systemic Discovery of New Genes
FILE REFERENCE: 032796-090
CURRENT APPLICATION NUMBER: US/10/083,357
CURRENT FILING DATE: 2002-02-27
NUMBER OF SEQ ID NOS: 1346
SEQ ID NO 1281
LENGTH: 120
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-083-357-1281

Query Match 2.9%; Score 6; DB 9; Length 120;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 DROFAP 69
DB 42 DROFAP 47

RESULT 32
US-10-050-704-107
Sequence 107, Application US/10050704
Publication No. US20030050442A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 62 Human Secreted Proteins
FILE REFERENCE: P2039P1
CURRENT APPLICATION NUMBER: US/10/050,704
CURRENT FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: 09/684,524
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: PCT/US00/08979
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/128,693
PRIOR FILING DATE: 1999-04-09
PRIOR APPLICATION NUMBER: 60/130,991
PRIOR FILING DATE: 1999-04-26
NUMBER OF SEQ ID NOS: 344
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 107
LENGTH: 136
TYPE: PRT
ORGANISM: Homo sapiens
US-10-050-704-107

Query Match 2.9%; Score 6; DB 9; Length 136;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 SALERA 176
DB 41 SALERA 46

RESULT 33
US-09-852-797-97
Sequence 97, Application US/09852797
Patent No. US20020172994A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P2
CURRENT APPLICATION NUMBER: US/09/852,797
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/040,710
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/050,934
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,100
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,357
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,189
PRIOR FILING DATE: 1997-05-30

PRIOR APPLICATION NUMBER: 60/057,765
PRIOR FILING DATE: 1997-09-05
PRIOR APPLICATION NUMBER: 60/048,970
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/068,368
PRIOR FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 97
LENGTH: 137
TYPE: PRT
ORGANISM: Homo sapiens
US-09-852-797-97

Query Match
Best Local Similarity 2.9%; Score 6; DB 9; Length 137;
Matches 6; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 IAISSL 45
Db 68 IAISSL 73

RESULT 34
US-09-853-161-97
Sequence 97, Application US/09853161
Patent No. US20020076756A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P3
CURRENT APPLICATION NUMBER: US/09/853,161
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/040,710
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/050,934
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,100
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,357
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,189
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/057,765
PRIOR FILING DATE: 1997-09-05
PRIOR APPLICATION NUMBER: 60/048,970
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/068,368
PRIOR FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 97
LENGTH: 137
TYPE: PRT
ORGANISM: Homo sapiens
US-09-853-161-97

Query Match
Best Local Similarity 2.9%; Score 6; DB 10; Length 137;
Matches 6; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 IAISSL 45
Db 68 IAISSL 73

RESULT 35
US-09-852-659A-97
Sequence 97, Application US/09852659A
Patent No. US20020077287A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P4
CURRENT APPLICATION NUMBER: US/09/852,659A
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/040,710
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/050,934
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,100
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,357
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,189
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/057,765
PRIOR FILING DATE: 1997-09-05
PRIOR APPLICATION NUMBER: 60/048,970
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/068,368
PRIOR FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 121
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 97
LENGTH: 137
TYPE: PRT
ORGANISM: Homo sapiens
US-09-852-659A-97

Query Match
Best Local Similarity 2.9%; Score 6; DB 10; Length 137;
Matches 6; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 IAISSL 45
Db 68 IAISSL 73

RESULT 36
US-09-925-297-799
Sequence 799, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 799
LENGTH: 138
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:

```
NAME/KEY: SITE
LOCATION: (126)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-799
```

```
Query Match
Best Local Similarity 2.9%; Score 6; DB 10; Length 138;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 121 LSGALI 126
Db 72 LSGALI 77
```

```
RESULT 37
US-10-156-761-11438
Sequence 11438, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11438
LENGTH: 152
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-11438
```

```
Query Match
Best Local Similarity 2.9%; Score 6; DB 9; Length 152;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 15 LVLSAC 20
Db 18 LVLSAC 23
```

```
RESULT 38
US-10-156-761-12938
Sequence 12938, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 12938
LENGTH: 152
```

```
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-12938
```

```
Query Match
Best Local Similarity 2.9%; Score 6; DB 9; Length 152;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 137 YEGRVV 142
Db 94 YEGRVV 99
```

```
RESULT 39
US-09-981-286A-1
Sequence 1, Application US/09981286A
Publication No. US20020192799A1
GENERAL INFORMATION:
APPLICANT: Watowich, Stanley J.
APPLICANT: Weaver, Scott C.
APPLICANT: Davey, Robert A.
TITLE OF INVENTION: Drug Discovery Methods
FILE REFERENCE: 265,00260101
CURRENT APPLICATION NUMBER: US/09/981,286A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: US 60/240,187
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 157
TYPE: PRT
ORGANISM: VENEZUELAN EQUINE ENCEPHALITIS VIRUS
US-09-981-286A-1
```

```
Query Match
Best Local Similarity 2.9%; Score 6; DB 9; Length 157;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 139 GRVVAL 144
Db 117 GRVVAL 122
```

```
RESULT 40
US-09-852-797-63
Sequence 63, Application US/09852797
Patent No. US20020172994A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P2
CURRENT APPLICATION NUMBER: US/09/852,797
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/040,710
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/050,934
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,100
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,357
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,189
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/057,765
```

; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 63
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-797-63

Query Match 2.9%; Score 6; DB 9; Length 162;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 IAISLL 45
|||
Db 76 IAISLL 81

Search completed: July 8, 2003, 11:56:10
Job time : 78 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:43:00 ; Search time 40 Seconds
(without alignments)
490.286 Million cell updates/sec

Title: US-10-069-544-2
Perfect score: 204
Sequence: 1 MKIRVMPWMMAMGLVLSAC.....EIKSYDKLSDYK 204

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	4.9	228	2 E69190	conserved hypotet
2	10	4.9	379	2 E69210	conserved hypotet
3	8	3.9	212	2 T01795	hypothetical prote
4	8	3.9	550	2 B71441	probable myosin II
5	8	3.9	830	2 T41509	serine/chreonine-P
6	8	3.9	1252	2 A47213	beta-fodrin - huma
7	8	3.9	2364	1 A44159	spectrin beta-G ch
8	7	3.4	78	1 E69549	conserved hypotet
9	7	3.4	86	2 D64371	hypothetical prote
10	7	3.4	103	2 AD3217	conserved hypotet
11	7	3.4	111	2 F70422	hypothetical prote
12	7	3.4	149	2 D59097	hypothetical prote
13	7	3.4	164	2 C69842	multidrug resistan
14	7	3.4	173	2 S75578	hypothetical prote
15	7	3.4	193	2 AH2035	hypothetical prote
16	7	3.4	204	2 T01040	hypothetical prote
17	7	3.4	228	2 AE1763	the cell-division
18	7	3.4	239	2 T40148	hypothetical prote
19	7	3.4	239	2 A89997	hypothetical prote
20	7	3.4	249	2 D85693	unknown protein en
21	7	3.4	262	2 G95902	probable tritoflito
22	7	3.4	268	2 H64647	co-chaperone-curve
23	7	3.4	288	2 C71936	probable co-chaper
24	7	3.4	292	2 T19669	hypothetical prote
25	7	3.4	292	2 G88783	protein C33A12.16
26	7	3.4	292	2 H82214	conserved hypotet
27	7	3.4	301	2 AF3504	acetyl-CoA carboxy
28	7	3.4	313	2 AD3233	DNA invertase (imp
29	7	3.4	324	2 AB0767	probable reductase

30	7	3.4	330	1 S15303	probable CDP-6-deo
31	7	3.4	335	2 T39033	hypothetical prote
32	7	3.4	336	2 A87300	TPR domain protein
33	7	3.4	336	2 T15727	hypothetical prote
34	7	3.4	349	2 T08782	hypothetical prote
35	7	3.4	362	2 C83980	hypothetical prote
36	7	3.4	378	2 C64142	conserved hypotet
37	7	3.4	378	2 AB0484	probable glycerate
38	7	3.4	395	2 A95860	hypothetical prote
39	7	3.4	423	2 S46741	glutamate 5-kinase
40	7	3.4	426	2 A12302	seryl-tRNA synthet
41	7	3.4	430	2 S75313	serine-tRNA ligase
42	7	3.4	432	2 AC2146	periplasmic sugar-
43	7	3.4	446	2 S67437	Damage and replica
44	7	3.4	452	2 F90452	conserved hypotet
45	7	3.4	458	2 S74941	hypothetical prote
46	7	3.4	461	2 AC0827	probable sensor Ki
47	7	3.4	470	2 D69984	glycolate oxidase
48	7	3.4	479	1 T46576	phosphoprotein pho
49	7	3.4	496	2 F91056	probable 2-compone
50	7	3.4	496	2 B85901	probable 2-compone
51	7	3.4	496	2 C65033	Putative sensor-11
52	7	3.4	499	1 A55346	phosphoprotein pho
53	7	3.4	530	2 D96810	hypothetical prote
54	7	3.4	553	2 C75318	hypothetical prote
55	7	3.4	566	2 F90406	succinate dehydrog
56	7	3.4	597	2 F90481	conserved hypotet
57	7	3.4	742	2 S56337	hypothetical prote
58	7	3.4	742	2 C91265	probable vimentin
59	7	3.4	742	2 H86105	probable vimentin
60	7	3.4	743	2 D84545	probable salt-indu
61	7	3.4	794	2 T09643	glutamine-tRNA lig
62	7	3.4	806	2 T35640	probable sensor Ki
63	7	3.4	822	2 AB2507	hypothetical prote
64	7	3.4	868	2 AF2517	hypothetical prote
65	7	3.4	873	2 H84459	hypothetical prote
66	7	3.4	885	2 AD2077	hypothetical prote
67	7	3.4	1066	2 T06001	hypothetical prote
68	7	3.4	1076	2 A69409	carbamoyl-phosphat
69	7	3.4	1118	2 S75309	hypothetical prote
70	7	3.4	1193	2 AC3240	helicase, SNF2 fam
71	7	3.4	1212	2 T28157	erythrocyte membra
72	7	3.4	2264	2 I40884	cytoxin L - C1os
73	6	2.9	24	2 S42785	relaxin - baboon (
74	6	2.9	31	2 A95237	conotoxin Tx5.1 pr
75	6	2.9	62	2 B59147	hypothetical prote
76	6	2.9	62	2 AF2187	hypothetical prote
77	6	2.9	75	2 A54355	light harvesting C
78	6	2.9	80	2 E97949	hypothetical prote
79	6	2.9	81	2 AC2331	30S ribosomal prot
80	6	2.9	88	2 PC4238	heat shock protein
81	6	2.9	89	2 G48909	dopamine receptor
82	6	2.9	89	2 G70880	hypothetical prote
83	6	2.9	90	2 AC1030	hypothetical prote
84	6	2.9	92	2 AF1969	hypothetical prote
85	6	2.9	92	2 D69218	conserved hypotet
86	6	2.9	93	2 B90684	hypothetical prote
87	6	2.9	93	2 D83951	hypothetical prote
88	6	2.9	93	2 F85534	unknown (imported)
89	6	2.9	93	2 H64767	hypothetical prote
90	6	2.9	94	2 A49855	heat shock protein
91	6	2.9	94	2 JC1479	heat shock protein
92	6	2.9	94	2 A83720	class I heat-shock
93	6	2.9	97	2 G64454	hypothetical prote
94	6	2.9	98	2 B29846	probable activator
95	6	2.9	107	2 S63991	thioredoxin trxa -
96	6	2.9	107	2 A86882	hypothetical prote
97	6	2.9	108	2 S40089	ribosomal protein
98	6	2.9	110	2 T29368	hypothetical prote
99	6	2.9	114	2 H86625	hypothetical prote
100	6	2.9	121	2 D82062	hypothetical prote
101	6	2.9	121	2 C84353	hypothetical prote
102	6	2.9	124	2 F82205	chemotaxis protein

103	2.9	126	2	P98103	hypothetical prote	176	209	2	E71897	molybdopterin-guan
104	2.9	127	2	E84058	hypothetical prote	177	210	2	G96832	hypothetical prote
105	2.9	129	2	B70832	hypothetical prote	178	212	2	UA0152	glycine chain A7
106	2.9	131	2	P00059	T-cell receptor be	179	212	2	F91002	probable prophage
107	2.9	132	2	A12337	acetyl-coenzyme A	180	213	2	T42080	superoxide dismuta
108	2.9	133	2	P85917	hypothetical prote	181	213	2	H95174	amino acid ABC tra
109	2.9	136	2	T12140	legumin - fava bea	182	213	2	C70555	hypothetical prote
110	2.9	137	2	F70659	hypothetical prote	183	213	2	C71088	hypothetical prote
111	2.9	139	2	A40001	ribose permease [I	184	214	2	A35063	hypothetical prote
112	2.9	141	2	A24338	hemoglobin alpha-T	185	215	2	F97261	response regulator
113	2.9	141	2	F81433	probable heme-bind	186	215	2	E85847	hypothetical prote
114	2.9	142	1	S03581	riboseosomal protei	187	215	2	C87522	hypothetical prote
115	2.9	142	2	A28538	hemoglobin alpha c	188	215	2	C37390	transcription regu
116	2.9	143	2	E84415	30S ribosomal prot	189	216	2	D86865	transfer protein T
117	2.9	144	2	E75059	hypothetical prote	190	216	2	AG1173	ribulose-phosphate
118	2.9	149	2	AF3343	hypothetical cytos	191	216	2	AH1530	hypothetical prote
119	2.9	151	2	DB3716	hypothetical prote	192	218	2	E95880	hypothetical prote
120	2.9	151	2	E83550	hypothetical prote	193	220	2	DMBYH	probable transcrip
121	2.9	154	2	JC4588	positive regulator	194	220	2	E82381	imidazoleglycerol-
122	2.9	155	2	T48503	RNA-binding protei	195	220	2	C87707	hypothetical prote
123	2.9	155	2	T48503	hypothetical prote	196	221	2	H72466	conserved hypotet
124	2.9	155	2	AB1226	hypothetical prote	197	221	2	D83213	conserved hypotet
125	2.9	155	2	AD2253	hypothetical prote	198	222	1	K1EYA	hypothetical prote
126	2.9	156	2	E83233	conserved hypotet	199	222	2	F95367	adenylate kinase (
127	2.9	158	2	C63064	hypothetical prote	200	224	2	S69635	hypothetical prote
128	2.9	161	2	A12064	hypothetical prote	201	225	2	C87595	hypothetical prote
129	2.9	162	2	P85928	probable response	202	226	2	D81001	isochorismatase fa
130	2.9	163	2	A43259	19K protein - Ente	203	228	2	G70902	conserved hypotet
131	2.9	164	2	S27650	2,3,4,5-tetrahydro	204	228	2	H71682	probable lipit prot
132	2.9	165	2	C42465	beta protein homol	205	228	2	AH3539	hypothetical prote
133	2.9	165	2	I40515	spai protein - Bac	206	232	1	S31235	imidazoleglycerol-
134	2.9	166	2	A40783	18-2K protein - ph	207	232	2	B88041	hypothetical prote
135	2.9	167	1	S32716	ribosome releasing	208	232	2	C85585	hypothetical prote
136	2.9	167	2	T44291	biotin carboxyl ca	209	232	2	B97035	unknown protein en
137	2.9	168	2	G69217	hypothetical prote	210	234	2	G85098	hypothetical prote
138	2.9	170	2	G69217	molybdenum cofacto	211	234	2	D70829	H+-transporting AT
139	2.9	170	2	T02408	hypothetical prote	212	236	2	AF5214	probable regulator
140	2.9	173	2	AB2205	hypothetical prote	213	236	2	AR2214	hypothetical prote
141	2.9	177	2	DB3727	RNA polymerase ECP	214	237	2	AR2196	phosphoribosylam
142	2.9	179	1	F64006	hypothetical prote	215	237	2	D40595	heme oxygenase [lm
143	2.9	179	2	D64709	hypothetical prote	216	237	2	B64643	hypothetical prote
144	2.9	181	2	D97452	hypothetical prote	217	238	2	T46505	hypothetical prote
145	2.9	181	2	AF2670	conserved hypotet	218				

249	6	2.9	254	2	S30595	alcohol dehydrogen	322	6	2.9	287	2	A83423	hypothetical prote
250	6	2.9	254	2	C40731	alcohol dehydrogen	323	6	2.9	288	2	S37687	heme oxygenase (de
251	6	2.9	254	2	T25470	hypothetical prote	324	6	2.9	288	2	F95123	conserved hypothet
252	6	2.9	254	2	H86458	29.0K hypothetical	325	6	2.9	288	2	H97993	hypothetical prote
253	6	2.9	256	1	DEPFA	alcohol dehydrogen	326	6	2.9	289	2	H75551	hypothetical prote
254	6	2.9	256	1	S07439	alcohol dehydrogen	327	6	2.9	289	2	T48108	hypothetical prote
255	6	2.9	256	1	S18273	alcohol dehydrogen	328	6	2.9	290	2	JC2097	legumin type B alp
256	6	2.9	256	1	S09633	alcohol dehydrogen	329	6	2.9	290	2	E72676	hypothetical prote
257	6	2.9	256	1	S09634	alcohol dehydrogen	330	6	2.9	291	2	T21475	hypothetical prote
258	6	2.9	256	1	S20713	alcohol dehydrogen	331	6	2.9	292	2	T35546	hypothetical prote
259	6	2.9	256	2	S18281	alcohol dehydrogen	332	6	2.9	293	2	T31840	hypothetical prote
260	6	2.9	256	2	S18282	alcohol dehydrogen	333	6	2.9	296	2	E96994	transcription regu
261	6	2.9	256	2	S18284	alcohol dehydrogen	334	6	2.9	297	2	H98302	hypothetical prote
262	6	2.9	256	2	S18287	alcohol dehydrogen	335	6	2.9	297	2	AF2980	conserved hypothet
263	6	2.9	256	2	S20718	alcohol dehydrogen	336	6	2.9	299	2	A82063	uroporphyrin-III C
264	6	2.9	256	2	S18377	alcohol dehydrogen	337	6	2.9	301	2	F70037	spore coat polysac
265	6	2.9	256	2	S18378	alcohol dehydrogen	338	6	2.9	301	2	A70787	hypothetical prote
266	6	2.9	256	2	S18379	alcohol dehydrogen	339	6	2.9	301	2	E97236	ABC-type multidrug
267	6	2.9	256	2	S18280	alcohol dehydrogen	340	6	2.9	302	2	E64800	probable glutamate
268	6	2.9	257	2	C70715	hypothetical prote	341	6	2.9	302	2	A85566	probable periplasm
269	6	2.9	257	2	S65958	maut protein - Par	342	6	2.9	302	2	F90715	probable periplasm
270	6	2.9	257	2	T34731	probable gas vesic	343	6	2.9	304	2	A11303	dihydroorotase den
271	6	2.9	257	2	AC1290	hypothetical prote	344	6	2.9	304	2	A11675	dihydroorotase den
272	6	2.9	257	2	A11661	hypothetical prote	345	6	2.9	305	2	B87393	probable 3',5'-cyc
273	6	2.9	258	2	C87458	dimethyladenosine	346	6	2.9	305	2	AC0319	probable amino aci
274	6	2.9	258	2	B45710	BRRF1 gene homolog	347	6	2.9	305	2	A71324	conserved hypothet
275	6	2.9	260	2	H72686	probable dehydroge	348	6	2.9	305	2	AC1588	hypothetical prote
276	6	2.9	260	2	H71979	probable type II r	349	6	2.9	305	2	AH1651	hypothetical prote
277	6	2.9	260	2	AD2461	hypothetical prote	350	6	2.9	308	1	S76941	catbamate kinase (
278	6	2.9	263	2	F64113	feric enterobacti	351	6	2.9	308	1	AF0583	ABC transporter pe
279	6	2.9	264	4	PKECT5	kanamycin kinase (352	6	2.9	309	2	S25511	outer membrane pro
280	6	2.9	264	4	S41977	kanamycin kinase (353	6	2.9	309	2	S65977	YybS protein - Bac
281	6	2.9	266	1	S55397	probable methyltra	354	6	2.9	309	2	D97607	trypsinesterase (ac
282	6	2.9	267	2	A96549	hypothetical prote	355	6	2.9	310	2	H69465	hypothetical prote
283	6	2.9	269	2	G87134	enoyl-[ACP] reduct	356	6	2.9	310	2	A49558	phytoene synthase
284	6	2.9	269	2	G70710	enoyl-[acyl-carrie	357	6	2.9	311	2	H82966	probable transcrip
285	6	2.9	269	2	F71519	probable rRNA meth	358	6	2.9	312	2	D95111	dihydroorotate den
286	6	2.9	270	2	A86164	protein F15K9.14 (359	6	2.9	312	2	S18387	heme oxygenase - r
287	6	2.9	270	2	AH2143	hypothetical prote	360	6	2.9	312	2	B82339	heme oxygenase (de
288	6	2.9	270	2	T40852	mecZ protein - Cox	361	6	2.9	313	2	S21700	heme oxygenase (de
289	6	2.9	272	2	D90136	hypothetical prote	362	6	2.9	313	2	B98339	integral membrane
290	6	2.9	274	2	AD0535	hypothetical prote	363	6	2.9	314	2	T02964	cyclin A-type (clo
291	6	2.9	274	2	S33178	protein kinase MNK	364	6	2.9	314	2	T08675	hypothetical prote
292	6	2.9	275	2	T43004	hypothetical prote	365	6	2.9	314	2	AH0516	probable activator
293	6	2.9	276	2	E71985	3-deoxy-d-manno-oc	366	6	2.9	314	2	C81735	tRNA delta-2-isope
294	6	2.9	276	2	C64520	3-deoxy-d-manno-oc	367	6	2.9	314	2	AF2637	hypothetical prote
295	6	2.9	278	2	B83137	hypothetical prote	368	6	2.9	315	1	A35199	heme oxygenase (de
296	6	2.9	278	2	A13633	hypothetical prote	369	6	2.9	315	2	JC5149	heme oxygenase (de
297	6	2.9	279	2	AB0873	hypothetical prote	370	6	2.9	315	2	A26044	prephenate dehydra
298	6	2.9	280	2	H83623	probable chemotaxi	371	6	2.9	316	2	I60119	heme oxygenase (de
299	6	2.9	280	2	T48244	hypothetical prote	372	6	2.9	316	2	E83136	probable transcrip
300	6	2.9	280	2	AC3579	transcription regu	373	6	2.9	316	2	B82485	iron(III) ABC tran
301	6	2.9	281	1	PNSAP	beta-lactamase (EC	374	6	2.9	317	2	G84414	homoserine dehydro
302	6	2.9	281	1	C95932	probable sugar upt	375	6	2.9	318	2	B36972	and 5'-region hypo
303	6	2.9	283	1	C47755	pectic enzyme secr	376	6	2.9	318	2	S33433	hypothetical prote
304	6	2.9	284	2	S06984	nitrogenase (EC 1.	377	6	2.9	321	2	D96035	probable transpos
305	6	2.9	284	2	T44267	hypothetical prote	378	6	2.9	321	2	E96025	probable transpos
306	6	2.9	284	2	T08547	hypothetical prote	379	6	2.9	322	2	T21478	hypothetical prote
307	6	2.9	284	2	C82181	conserved hypothet	380	6	2.9	322	2	S25513	outer membrane pro
308	6	2.9	285	2	B83257	rRNA-pseudouridine	381	6	2.9	323	2	S25515	outer membrane pro
309	6	2.9	285	2	A64217	ribosomal protein	382	6	2.9	323	2	S25514	outer membrane pro
310	6	2.9	285	2	C25242	tropomyosin, exon	383	6	2.9	323	2	A44801	albumin-binding pr
311	6	2.9	285	2	A25561	tropomyosin II, mu	384	6	2.9	325	2	T46873	electron transfer
312	6	2.9	286	2	B25242	tropomyosin, exon	385	6	2.9	326	2	AF2082	iron(III) diglutar
313	6	2.9	286	2	A25242	tropomyosin, exon	386	6	2.9	329	1	CYGPZ	zeta-crystallin /
314	6	2.9	286	2	T03434	probable transport	387	6	2.9	329	1	S07577	legumin storage pr
315	6	2.9	286	2	AE3245	hypothetical prote	388	6	2.9	330	2	S75139	glycerol-3-phospha
316	6	2.9	286	2	AH2943	hypothetical prote	389	6	2.9	330	2	B97980	dihydroorotase oxi
317	6	2.9	286	2	T02651	hypothetical prote	390	6	2.9	330	2	D83360	cell division prot
318	6	2.9	286	2	T16532	hypothetical prote	391	6	2.9	331	2	C90486	ABC transporter, A
319	6	2.9	286	2	S72712	Leptb170_F1_44 pro	392	6	2.9	331	2	T01838	hypothetical prote
320	6	2.9	286	2	B83799	hypothetical prote	393	6	2.9	333	2	S60615	RNA-directed RNA p
321	6	2.9	286	2	AF2829	hydrolyase (importe	394	6	2.9	333	2	F83331	probable transcrip

395	6	2.9	334	2	C64083	tryptophan-tRNA 11
396	6	2.9	334	2	A12150	serine/threonine k
397	6	2.9	335	2	S07576	legumin storage pr
398	6	2.9	335	2	S07578	legumin storage pr
399	6	2.9	335	2	AC3160	NAD binding oxidor
400	6	2.9	336	2	B83255	probable aspartate
401	6	2.9	336	2	B71366	probable phosphate
402	6	2.9	336	2	H69282	DNA repair protei
403	6	2.9	337	2	G95822	probable glycosyl
404	6	2.9	337	2	T52259	nitrite (EC 3.5.
405	6	2.9	339	2	T52262	nitrite (EC 3.5.
406	6	2.9	339	2	S31969	nitrite (EC 3.5.
407	6	2.9	339	2	A71473	probable tRNA pyro
408	6	2.9	339	2	S77404	cell division limi
409	6	2.9	339	2	T25794	hypothetical prote
410	6	2.9	339	2	G82113	lipoprotein-34 NIP
411	6	2.9	340	1	G82113	gene V protein - p
412	6	2.9	340	1	T24615	hypothetical prote
413	6	2.9	340	1	T24615	hypothetical prote
414	6	2.9	340	1	T24615	hypothetical prote
415	6	2.9	343	2	B82377	moco protein limpo
416	6	2.9	343	2	B82377	spermidine/putresc
417	6	2.9	346	2	H81915	protein-glutamate
418	6	2.9	346	2	T49147	probable alcohol d
419	6	2.9	346	2	S22398	nitrite (EC 3.5.
420	6	2.9	346	2	AB0020	nitrite (EC 3.5.
421	6	2.9	346	2	E95258	tryptophan-tRNA 11
422	6	2.9	346	2	H98123	translation elonga
423	6	2.9	346	2	AG0302	elongation factor
424	6	2.9	347	2	AG0302	Lact family transc
425	6	2.9	348	2	H81186	probable geranylge
426	6	2.9	348	2	H81186	probable alcohol d
427	6	2.9	348	2	T12591	NADH dehydrogenas
428	6	2.9	349	2	T07669	cyclin al-type, mi
429	6	2.9	350	2	H75468	conserved hypotet
430	6	2.9	351	2	S00337	legumin B legk pre
431	6	2.9	351	2	E85925	hypothetical prote
432	6	2.9	352	1	D91080	hypothetical prote
433	6	2.9	352	1	K18E84	thymidine kinase (
434	6	2.9	352	2	T42582	thymidine kinase (
435	6	2.9	353	2	A69743	ATP-binding Mrp-11
436	6	2.9	353	2	B89908	threonine synthase
437	6	2.9	355	1	LKCH	proteoglycan link
438	6	2.9	355	2	T01737	hypothetical prote
439	6	2.9	356	2	C70025	multidrug-efflux t
440	6	2.9	356	2	C84203	glucose dehydrogen
441	6	2.9	357	2	T34133	steroid/hydroid/re
442	6	2.9	357	2	A88204	hypothetical prote
443	6	2.9	357	2	C86050	hypothetical prote
444	6	2.9	360	2	S17456	outer membrane pro
445	6	2.9	361	2	I40347	recombination prot
446	6	2.9	363	2	AG0346	putative ABC-trans
447	6	2.9	364	2	T50475	dopamine D1 recept
448	6	2.9	364	2	G71410	probable protein k
449	6	2.9	364	2	T44672	H+-transporting At
450	6	2.9	365	2	G70364	conserved hypotet
451	6	2.9	365	2	C72308	conserved hypotet
452	6	2.9	365	2	T29234	hypothetical prote
453	6	2.9	366	1	DEB97C	3-isopropylmalate
454	6	2.9	366	1	SZB85E	stage V sporulatio
455	6	2.9	366	2	U00513	phenylalanine dehy
456	6	2.9	367	2	C81713	DNA polymerase III
457	6	2.9	368	2	S01431	modulation protein
458	6	2.9	368	2	F84209	hypothetical prote
459	6	2.9	369	2	G72730	hypothetical prote
460	6	2.9	370	1	LNRRB	pulmonary surfacta
461	6	2.9	370	2	H70583	phosphate-binding
462	6	2.9	371	2	I40358	N-acyl-L-ornithin
463	6	2.9	372	2	S68413	phosphoenolpyruvat
464	6	2.9	372	2	AH0703	probable membrane
465	6	2.9	373	2	S39722	spore coat polysac
466	6	2.9	374	2	S65077	1,3-beta-glucanase
467	6	2.9	374	2	E87596	hypothetical prote
	6	2.9	374	2	E91250	probable proteins
	6	2.9	374	2	F98308	probable peptidic
468	6	2.9	374	2	AE2974	acetylornithine de
469	6	2.9	375	2	B83547	citrate synthase 2
470	6	2.9	376	2	G70800	probable ABC trans
471	6	2.9	376	2	B65081	hypothetical prote
472	6	2.9	378	2	G91107	probable oxidase I
473	6	2.9	378	2	B85593	probable oxidase I
474	6	2.9	378	2	AE3350	reca protein limpo
475	6	2.9	379	2	JC5944	muconate cyclisom
476	6	2.9	379	2	F84144	carboxymalate
477	6	2.9	380	2	A82244	response regulator
478	6	2.9	381	2	A82514	conserved hypotet
479	6	2.9	381	2	T46827	phenoxymethylate di
480	6	2.9	382	2	AC2328	type II activin re
481	6	2.9	382	2	B49193	hypothetical prote
482	6	2.9	382	2	E85082	hypothetical prote
483	6	2.9	382	2	T14186	hypothetical prote
484	6	2.9	384	1	I38890	dual specificity p
485	6	2.9	384	1	T02868	cyclin A-type (Cio
486	6	2.9	384	2	A35003	bombesin/gastrin-r
487	6	2.9	384	2	I57682	bombesin/GRP rece
488	6	2.9	384	2	A41007	gastrin-releasing
489	6	2.9	384	2	A69622	ferrichrome ABC tr
490	6	2.9	385	2	T13692	hypothetical prote
491	6	2.9	385	2	F87684	hypothetical prote
492	6	2.9	386	2	E72207	GRP cyclohydrolase
493	6	2.9	388	2	AD1378	cell division prot
494	6	2.9	389	2	AE1747	cell division prot
495	6	2.9	390	2	JH0374	bombesin receptor,
496	6	2.9	390	2	B41007	bombesin receptor,
497	6	2.9	393	2	B96780	alcohol dehydrogen
498	6	2.9	393	2	C69292	dihydroxyphenylam
499	6	2.9	395	2	D64379	hypothetical prote
500	6	2.9	397	2	T08716	hypothetical prote
501	6	2.9	399	2	S07630	protein-tyrosine-p
502	6	2.9	399	2	F83484	probable MFS trans
503	6	2.9	399	2	T35803	probable ABC-trans
504	6	2.9	400	2	F75275	chromate transport
505	6	2.9	402	2	T23701	hypothetical prote
506	6	2.9	402	2	T40066	probable vacuolar
507	6	2.9	402	2	S46641	probable membrane
508	6	2.9	403	2	T39697	DNAJ protein - fis
509	6	2.9	403	2	AC2271	presorcin-6y-depen
510	6	2.9	404	2	A75553	Grp cyclohydrolase
511	6	2.9	404	2	B84745	probable RNA-bind
512	6	2.9	404	2	B72768	hypothetical prote
513	6	2.9	404	2	H70620	hypothetical prote
514	6	2.9	404	2	T12151	hypothetical prote
515	6	2.9	404	2	D59945	probable integral
516	6	2.9	405	2	T01218	conserved hypotet
517	6	2.9	406	2	B89942	molymdenum cofacto
518	6	2.9	407	2	H83155	hypothetical prote
519	6	2.9	407	2	D85643	probable MYST-fam
520	6	2.9	407	2	T37865	probable protein - hu
521	6	2.9	409	2	S36113	lysine protein - hu
522	6	2.9	409	2	AG0205	probable alanine r
523	6	2.9	410	2	S68307	phycocyanin synthase
524	6	2.9	410	2	S48052	beta-lactam-activat
525	6	2.9	410	2	G95071	beta-lactam resist
526	6	2.9	410	2	E97939	alanine adding enz
527	6	2.9	411	2	G97802	cytosine-tRNA liga
528	6	2.9	412	2	C23783	hypothetical prote
529	6	2.9	413	2	T40427	dhaf related prote
530	6	2.9	413	2	B82760	cell division prot
531	6	2.9	414	2	E84779	hypothetical prote
532	6	2.9	414	2	T30829	hypothetical prote
533	6	2.9	414	2	E83335	RND multidrug effl
534	6	2.9	415	2	A56042	mitogen-activated
535	6	2.9	415	2	G64510	hypothetical prote
536	6	2.9	416	2	E71559	probable DNA pol I
537	6	2.9	416	2	C95973	probable exported
538	6	2.9	417	2	H82315	serine transporter
539	6	2.9	417	2	E70207	antigen SI - Lyme
540	6	2.9	417	2	D82069	conserved hypotet

541	6	2.9	419	2	G82009	SUN homolog	614	6	2.9	466	2	T44650	capsular polysacch
542	6	2.9	419	2	H96601	hypothetical prote	615	6	2.9	467	2	A99494	thermostable carbo
543	6	2.9	419	2	B81236	16S RNA methyltran	616	6	2.9	467	2	I56896	gene fli-2 protein
544	6	2.9	421	2	S73010	hypothetical prote	617	6	2.9	467	2	AD2449	hypothetical prote
545	6	2.9	421	2	D69981	conserved hypotet	618	6	2.9	468	2	H87044	hypothetical prote
546	6	2.9	422	2	A99783	probable tail tip	619	6	2.9	469	2	T45201	hypothetical prote
547	6	2.9	423	2	S61868	acetylornithine tr	620	6	2.9	470	2	A65168	hypothetical 49.9
548	6	2.9	423	2	B64079	probable amidohydr	621	6	2.9	470	2	AC1861	hypothetical prote
549	6	2.9	423	2	AH1407	PTS system galacti	622	6	2.9	471	2	AF3324	UDP-N-acetylmuram
550	6	2.9	423	2	AH1783	PTS system galacti	623	6	2.9	471	2	A45616	antigenic protein
551	6	2.9	425	2	D70902	GTP cyclohydrolase	624	6	2.9	472	1	B65098	hexuronate transpo
552	6	2.9	425	2	C70327	serine-CRNA ligase	625	6	2.9	472	2	G91125	transport of hexur
553	6	2.9	425	2	D72261	serine-CRNA ligase	626	6	2.9	472	2	F85970	hypothetical prote
554	6	2.9	425	2	H83652	seryl-tRNA synthet	627	6	2.9	474	2	T21771	hypothetical prote
555	6	2.9	425	2	A71639	serine-CRNA ligase	628	6	2.9	475	2	G81227	probable amino-act
556	6	2.9	425	2	H97851	serine-CRNA ligase	629	6	2.9	478	2	C81103	alginase O-acetyla
557	6	2.9	426	2	D83890	sugar transport sy	630	6	2.9	478	2	H81838	probable polysacch
558	6	2.9	427	2	AB1418	seryl-tRNA synthet	631	6	2.9	482	2	T49079	serine-type carbox
559	6	2.9	427	2	AD1793	seryl-tRNA synthet	632	6	2.9	482	2	T48337	hypothetical prote
560	6	2.9	428	2	G83330	outer membrane pro	633	6	2.9	483	2	T23312	hypothetical prote
561	6	2.9	430	2	A87541	major facilitator	634	6	2.9	484	2	A24942	legumin B4 precurs
562	6	2.9	430	2	A87708	hypothetical prote	635	6	2.9	484	2	C82426	codyric acid synth
563	6	2.9	431	1	C69689	response regulator	636	6	2.9	484	2	F71061	hypothetical prote
564	6	2.9	431	2	S45038	protein disulfide-	637	6	2.9	484	2	E86416	unknown protein, 3
565	6	2.9	431	2	T02496	probable protein k	638	6	2.9	484	2	T26393	hypothetical prote
566	6	2.9	432	2	S49980	glutamate-5-semial	639	6	2.9	485	2	S44268	legumin B precursor
567	6	2.9	432	2	T21232	hypothetical prote	640	6	2.9	487	1	DVRTD1	dopamine receptor
568	6	2.9	434	2	AG0071	Exu transport pro	641	6	2.9	487	2	B84472	probable serine ca
569	6	2.9	434	2	D82237	inosine-guanosine	642	6	2.9	487	2	T49080	serine-type carbox
570	6	2.9	436	2	B82431	regulatory protein	643	6	2.9	490	2	G95900	probable carboxydr
571	6	2.9	437	2	S67305	phosphopyruvate hy	644	6	2.9	493	2	A33809	carillage matrix p
572	6	2.9	437	2	S69881	phosphopyruvate hy	645	6	2.9	495	2	T36043	probable integrase
573	6	2.9	438	2	B82077	outer membrane pro	646	6	2.9	496	2	S51597	signal recognition
574	6	2.9	439	1	XNHUO	ornithine-oxo-acid	647	6	2.9	496	2	F84664	hypothetical prote
575	6	2.9	439	1	XNMSO	ornithine-oxo-acid	648	6	2.9	497	2	T06185	signal recognition
576	6	2.9	439	1	XNRTO	ornithine-oxo-acid	649	6	2.9	497	2	T06186	signal recognition
577	6	2.9	439	1	S19656	protein disulfide-	650	6	2.9	497	2	D97264	galactose-1-phosph
578	6	2.9	439	2	F86884	preprotein translo	651	6	2.9	497	2	F86150	F2288.2 protein -
579	6	2.9	439	2	F75306	probable sun prote	652	6	2.9	498	1	S52570	phosphoprotein pho
580	6	2.9	439	2	AC3333	replication initiat	653	6	2.9	500	2	S26688	legumin K - garden
581	6	2.9	440	1	S34379	ferredoxin-NADP re	654	6	2.9	501	2	S74341	hypothetical prote
582	6	2.9	440	2	AB2321	ferredoxin-NADP (+)	655	6	2.9	502	2	T35356	probable regulator
583	6	2.9	440	2	UC4369	p5 protein precurs	656	6	2.9	502	2	AG3641	histidine-tRNA lig
584	6	2.9	440	2	T52320	10-deacetylthiobact	657	6	2.9	503	2	S00336	legumin B legu pre
585	6	2.9	441	2	A44815	cellulase (EC 3.2.	658	6	2.9	503	2	H72359	hypothetical prote
586	6	2.9	441	2	D95124	glycosyl transfera	659	6	2.9	504	2	B40829	hypothetical prote
587	6	2.9	441	2	P97994	conserved hypotet	660	6	2.9	504	2	S18679	capsid protein - g
588	6	2.9	443	2	G81471	biotin carboxylase	661	6	2.9	505	1	D70703	ctld homolog Rv231
589	6	2.9	443	2	D72383	NADH oxidase - The	662	6	2.9	505	2	T10896	cytochrome P450 (E
590	6	2.9	444	2	H71243	probable helicase	663	6	2.9	508	2	T01937	hypothetical prote
591	6	2.9	446	2	I47217	dopamine receptor	664	6	2.9	510	1	A4635	activin receptor S
592	6	2.9	446	2	A75209	DNA repair protein	665	6	2.9	510	1	A56926	activin receptor I
593	6	2.9	447	2	T39170	probable cis-mucon	666	6	2.9	511	2	UC5880	activin receptor I
594	6	2.9	447	2	T17299	hypothetical prote	667	6	2.9	511	2	A56750	archaen - human
595	6	2.9	448	2	E81811	hypothetical prote	668	6	2.9	512	2	D40829	activin receptor I
596	6	2.9	450	2	A55886	dopamine receptor	669	6	2.9	512	2	I37134	activin type II re
597	6	2.9	450	2	B82957	probable aminotran	670	6	2.9	513	2	D83711	alronate oxidore
598	6	2.9	450	2	AI0119	conserved hypotet	671	6	2.9	513	2	SYCHLZ	5-aminolevullinate
599	6	2.9	450	2	AI0178	conserved hypotet	672	6	2.9	513	1	UQ1486	activin receptor I
600	6	2.9	451	2	I51659	dopamine DIA recep	673	6	2.9	513	2	D72753	probable xylose
601	6	2.9	451	2	AH0827	probable PTS syste	674	6	2.9	513	2	S23089	activin receptor t
602	6	2.9	458	2	E75397	phosphoribosylamin	675	6	2.9	513	2	A39896	activin receptor p
603	6	2.9	459	2	A56849	dopamine receptor-	676	6	2.9	513	2	A40193	type II activin re
604	6	2.9	459	2	S76122	hypothetical prote	677	6	2.9	513	2	S27258	activin receptor t
605	6	2.9	460	2	E83517	probable aminotran	678	6	2.9	513	2	UQ1484	activin receptor t
606	6	2.9	461	2	G70379	cell division prot	679	6	2.9	513	2	I45850	activin receptor t
607	6	2.9	462	2	AH3621	chloride channel p	680	6	2.9	513	2	T43867	phosphatase transp
608	6	2.9	462	2	T08199	hypothetical prote	681	6	2.9	514	2	E71974	flagellin B - Heii
609	6	2.9	463	2	AH1566	galactosamine-cont	682	6	2.9	514	2	C64534	flagellin B - Heii
610	6	2.9	463	2	T32111	acid phosphatase (683	6	2.9	515	2	C88487	protein C13B9.3 (i
611	6	2.9	464	1	B59300	glutamy-tRNA amid	684	6	2.9	516	2	S38486	Na+-dependent symp
612	6	2.9	465	2	A63102	probable amidase p	685	6	2.9	517	2	A45175	variant surface gl
613	6	2.9	465	2	E98184		686	6	2.9	517	2	C35480	variant surface gl

687	2.9	518	2	AG0784	run protein [impor
688	2.9	518	2	G71646	probable proteogly
689	2.9	519	2	E72021	60k chaperonin, pr
690	2.9	519	2	H86602	heat shock protein
691	2.9	520	2	T18124	probable capsid pr
692	2.9	521	2	AG3475	RNA polymerase sig
693	2.9	521	2	T08569	glucose-1-phosphat
694	2.9	521	2	D83399	probable ATP-bind
695	2.9	522	2	T15119	hypothetical prote
696	2.9	524	2	B97864	hypothetical prote
697	2.9	526	2	T00440	hypothetical prote
698	2.9	526	2	T00804	hypothetical prote
699	2.9	527	2	T05548	hypothetical prote
700	2.9	528	2	C40829	activin receptor i
701	2.9	528	2	C87663	conserved hypotet
702	2.9	529	2	G70140	oligopeptide ABC t
703	2.9	531	2	B83371	conserved hypotet
704	2.9	533	2	B84858	phosphoprotein pho
705	2.9	533	2	T52063	ran GTPase-activat
706	2.9	535	2	D70429	ribosomal protein
707	2.9	536	2	A40839	activin receptor i
708	2.9	539	1	S49435	cytochrome-c oxida
709	2.9	539	2	A84534	hypothetical prote
710	2.9	544	2	S41090	triacylglycerol 11
711	2.9	544	2	S41091	triacylglycerol 11
712	2.9	544	2	S41092	triacylglycerol 11
713	2.9	544	2	S59588	triacylglycerol 11
714	2.9	544	2	B97156	flagellar basal bo
715	2.9	545	2	A38447	oligopeptide ABC t
716	2.9	548	2	B87443	cytochrome-c oxida
717	2.9	548	2	F69492	hypothetical prote
718	2.9	549	2	D90477	quinol oxidase-2,
719	2.9	549	2	G71606	hypothetical prote
720	2.9	550	2	T34230	hypothetical prote
721	2.9	553	2	A89830	arginyl-CRNA synth
722	2.9	553	2	A49364	59 protein, brain
723	2.9	555	2	A87458	sensor histidine k
724	2.9	556	2	B82141	ribosomal protein
725	2.9	557	2	T16815	hypothetical prote
726	2.9	557	2	S23429	protein kinase ERK
727	2.9	557	2	T50788	ubiquitin specific
728	2.9	558	2	A98159	translocated intim
729	2.9	558	2	E86045	probable transloca
730	2.9	559	2	C75286	hypothetical prote
731	2.9	559	2	B95130	site-specific reco
732	2.9	559	2	F97989	site-specific reco
733	2.9	560	2	S11004	glycinin G4 precu
734	2.9	560	2	T39609	transcription regu
735	2.9	561	2	T36084	hypothetical prote
736	2.9	562	2	FMSYG5	glycinin chain A5A
737	2.9	562	2	S20946	glycinin G4A precu
738	2.9	563	1	ACGUGC	triacylglycerol 11
739	2.9	563	2	S54802	glycinin A5A4B3 ch
740	2.9	564	2	S37241	legumin B - fava b
741	2.9	565	2	AC0184	malate dehydrogena
742	2.9	568	2	I58106	gene DMR-N9 protei
743	2.9	570	2	AH2765	cytochrome-c oxida
744	2.9	571	2	T31627	hypothetical prote
745	2.9	571	2	H89866	hypothetical prote
746	2.9	572	2	H86355	hypothetical prote
747	2.9	575	2	H64137	phosphotransferase
748	2.9	575	2	F83904	hypothetical prote
749	2.9	579	2	AF1788	DNA polymerase III
750	2.9	581	1	NDECE2	colicin E2 (EC 3.1
751	2.9	581	2	B84715	hypothetical prote
752	2.9	582	2	A32034	choline kinase (EC
753	2.9	583	2	T04531	nine-cis-epoxycaro
754	2.9	585	2	F90013	conserved hypotet
755	2.9	588	2	T45564	hypothetical prote
756	2.9	590	2	H84388	excision nuclease
757	2.9	591	2	B83457	glyoxylate carboxyl
758	2.9	591	2	E83039	probable thiol-dis
759	2.9	592	1	IKBCCA	colicin A - Citrub
760	2.9	592	2	AC3442	acyl-CoA dehydroge
761	2.9	594	2	AH2936	methy1-accepting c
762	2.9	594	2	G98345	probable mcpA prot
763	2.9	598	2	B97546	cytochrome c oxida
764	2.9	599	2	A48863	limonene cyclase -
765	2.9	602	2	F71617	SERA antigen/papai
766	2.9	612	2	G83307	hypothetical prote
767	2.9	613	2	AC2962	hypothetical prote
768	2.9	616	2	A11180	amylase homolog 1
769	2.9	619	2	C70669	probable acyl-CoA
770	2.9	621	1	S59632	endo-1,4-beta-xyla
771	2.9	621	2	T20307	hypothetical prote
772	2.9	624	2	A84150	hypothetical prote
773	2.9	627	2	B84482	probable gag-prote
774	2.9	627	2	G81719	signal peptidase,
775	2.9	628	2	D70722	hypothetical prote
776	2.9	633	2	C98321	C4-dicarboxylate t
777	2.9	646	2	S72609	GTP-binding membra
778	2.9	653	2	T34356	hypothetical prote
779	2.9	654	2	T45142	hypothetical prote
780	2.9	657	2	S73946	squalene-hopene cy
781	2.9	657	2	S73428	exonuclease ABC c
782	2.9	659	2	S77658	probable lipoprote
783	2.9	664	2	A40588	hypothetical prote
784	2.9	666	2	D86905	NADH2 dehydrogenas
785	2.9	667	2	JC5889	ATP-dependent DNA
786	2.9	668	2	JC3263	OS-9 protein precu
787	2.9	669	2	T06702	protein kinase (EC
788	2.9	671	2	D38490	hypothetical prote
789	2.9	672	2	T20310	matutase-related p
790	2.9	673	2	AC0073	hypothetical prote
791	2.9	680	2	D86925	2,4-dienoyl-CoA re
792	2.9	681	2	S75795	probable acyl-CoA
793	2.9	682	2	T07024	NADH2 dehydrogenas
794	2.9	684	2	A53019	dnak-type molecula
795	2.9	684	2	G84730	collagen alpha 1(X
796	2.9	684	2	A82134	Metulor-like trans
797	2.9	687	2	F83100	metulor-accepting c
798	2.9	689	2	T29772	conserved hypotet
799	2.9	690	2	S35251	hypothetical prote
800	2.9	690	2	H84309	probable membrane
801	2.9	691	2	B81937	protophyrin IX
802	2.9	691	2	S39867	competence protein
803	2.9	691	2	G81167	competence protein
804	2.9	693	2	D87437	TonB-dependent rec
805	2.9	693	2	T23529	hypothetical prote
806	2.9	694	2	D84811	hypothetical prote
807	2.9	694	2	B84331	hypothetical prote
808	2.9	696	2	AE2300	NADH dehydrogenase
809	2.9	696	2	C65007	probable fatty oxi
810	2.9	714	2	H91031	probable enzyme 11
811	2.9	714	2	A85876	probable enzyme 23
812	2.9	714	2	F70170	chemotaxis histidil
813	2.9	715	2	T25233	hypothetical prote
814	2.9	717	2	T34514	hypothetical prote
815	2.9	719	2	A10769	protein-tyrosine k
816	2.9	721	2	A82941	hypothetical prote
817	2.9	722	2	B71728	probable peptidase
818	2.9	725	2	B82425	conserved hypotet.
819	2.9	727	2	B60191	transcription regu
820	2.9	732	2	A90987	hypothetical prote
821	2.9	732	2	D85832	hypothetical prote
822	2.9	732	2	C64972	probable ATPase -
823	2.9	733	1	S56767	replication licens
824	2.9	733	2	E95335	probable cation tr
825	2.9	733	2	T11668	tnal4 protein - fl
826	2.9	734	1	S09156	diacylglycerol kin
827	2.9	734	1	I38080	replication licens
828	2.9	735	2	S12969	diacylglycerol kin
829	2.9	740	2	F82614	conserved hypotet
830	2.9	747	2	AE2929	two component resp
831	2.9	750	1	COZPME	melt protein - fls
832	2.9	750	2	ABO708	catalase (EC 1.11.

833	6	2.9	750	2	T48804	hypothetical prote
834	6	2.9	751	2	A81816	nitric-oxide reduc
835	6	2.9	751	2	D81062	nitric oxide reduc
836	6	2.9	753	2	A39129	catalase (EC 1.11.
837	6	2.9	753	2	B85782	catalase, hydroper
838	6	2.9	753	2	F90933	catalase, HPII (imp
839	6	2.9	753	2	C71086	hypothetical prote
840	6	2.9	753	2	C89786	hypothetical prote
841	6	2.9	755	2	S42462	structural polypro
842	6	2.9	756	2	C87432	hypothetical prote
843	6	2.9	760	2	S62782	probable lipoprote
844	6	2.9	763	2	S46678	hypothetical prote
845	6	2.9	783	2	A98353	probable transcrip
846	6	2.9	786	2	H69980	single-strand DNA-
847	6	2.9	791	2	H72552	hypothetical prote
848	6	2.9	802	2	C86437	hypothetical prote
849	6	2.9	809	2	T32899	probable leukotrie
850	6	2.9	811	2	S57149	probable membrane
851	6	2.9	815	2	T00538	probable serine pr
852	6	2.9	815	2	T36671	probable helicase
853	6	2.9	816	2	T08978	serine proteinase
854	6	2.9	818	2	T15803	hypothetical prote
855	6	2.9	828	2	AH2443	hypothetical prote
856	6	2.9	831	2	S50163	nitrate reductase
857	6	2.9	837	2	B84612	hypothetical prote
858	6	2.9	839	2	T31659	hypothetical prote
859	6	2.9	840	2	C75313	probable proteinas
860	6	2.9	843	2	S38364	membrane alanyl am
861	6	2.9	844	2	AC1981	hypothetical prote
862	6	2.9	848	2	B84107	hypothetical prote
863	6	2.9	851	2	H84455	probable receptor-
864	6	2.9	851	2	T22696	hypothetical prote
865	6	2.9	851	2	AG2469	hypothetical prote
866	6	2.9	852	2	D72330	conserved hypothet
867	6	2.9	854	1	VCLJST	env polyprotein pr
868	6	2.9	854	1	WMBP22	gene 12 protein -
869	6	2.9	854	1	WMBP22	gene 12 protein -
870	6	2.9	855	2	T19405	hypothetical prote
871	6	2.9	861	2	H59091	hypothetical prote
872	6	2.9	861	2	T27164	hypothetical prote
873	6	2.9	862	2	S64821	probable membrane
874	6	2.9	865	1	DASYL1	lipoxigenase (EC 1
875	6	2.9	865	2	B90644	aconitase hydrazase
876	6	2.9	865	2	B85495	aconitase hydrazase
877	6	2.9	865	2	AE0522	aconitase hydratase
878	6	2.9	865	2	AI0414	aconitase hydratase
879	6	2.9	865	2	F64734	aconitase hydratase
880	6	2.9	867	2	AH0062	probable Clp ATPase
881	6	2.9	868	2	AE1953	hypothetical prote
882	6	2.9	877	2	S58824	probable membrane
883	6	2.9	882	2	H97479	aminopeptidase N (
884	6	2.9	882	2	AH2597	aminopeptidase N p
885	6	2.9	887	2	AG0535	serine/threonine p
886	6	2.9	888	2	A55318	dup leucine zipper
887	6	2.9	888	2	JC0539	Clp ATPase (import
888	6	2.9	891	2	AE0358	hypothetical prote
889	6	2.9	895	2	T49010	hypothetical prote
890	6	2.9	897	2	T06540	polyribonucleotide
891	6	2.9	901	2	G71286	probable pyruvate,
892	6	2.9	905	2	C70758	hypothetical prote
893	6	2.9	912	2	H90567	hypothetical prote
894	6	2.9	916	2	F71962	hypothetical prote
895	6	2.9	923	2	S44226	periodic tryptophan
896	6	2.9	927	2	T38518	ribonuclease II RN
897	6	2.9	933	2	T41122	nucleoporin - fibs
898	6	2.9	934	2	S75633	hypothetical prote
899	6	2.9	941	2	T49136	protein kinase-lik
900	6	2.9	943	2	S59317	DIP2 protein - Yea
901	6	2.9	963	2	F86739	teichoic acid bios
902	6	2.9	967	2	A64710	type III restricti
903	6	2.9	970	2	J00302	hypothetical 112k
904	6	2.9	979	2	J00894	p15 protein - Myc
905	6	2.9	979	2	B90601	hypothetical prote
906	6	2.9	960	2	G75523	probable cell divi
907	6	2.9	985	2	D82776	pyruvate dehydroge
908	6	2.9	988	2	F86316	protein T10022.13
909	6	2.9	1006	2	S74992	hypothetical prote
910	6	2.9	1009	2	AF2033	hypothetical prote
911	6	2.9	1013	2	B70841	probable heliz prot
912	6	2.9	1015	1	DJBE2L	DNA-directed DNA p
913	6	2.9	1020	1	OFHGH	neurofilament trip
914	6	2.9	1027	2	A56533	chromatin remodell
915	6	2.9	1028	2	S41749	myosin heavy chain
916	6	2.9	1028	2	S37146	myosin I heavy cha
917	6	2.9	1035	2	I58409	integrin alpha-9 c
918	6	2.9	1041	2	T29010	hypothetical prote
919	6	2.9	1042	2	T29307	hypothetical prote
920	6	2.9	1042	2	S41705	Evil protein - hum
921	6	2.9	1042	2	A31591	Transcription regu
922	6	2.9	1051	2	A60191	oncogene Evi-1 - h
923	6	2.9	1058	2	T30580	p-type ATPase - 81
924	6	2.9	1058	2	B95292	probable adenylate
925	6	2.9	1060	2	S33641	homeotic protein 2
926	6	2.9	1067	2	B81851	probable drug effl
927	6	2.9	1067	2	B81851	multiple transfera
928	6	2.9	1077	2	S66842	hypothetical prote
929	6	2.9	1095	2	S76044	hypothetical prote
930	6	2.9	1096	2	T13802	centrosome-associa
931	6	2.9	1102	2	A32247	virg protein - Shi
932	6	2.9	1116	2	I54378	gene X104 protein
933	6	2.9	1129	2	D84022	ATP-dependent nucl
934	6	2.9	1131	2	S34213	H+-exporting ATPas
935	6	2.9	1145	2	S37136	structural polypro
936	6	2.9	1155	2	G64332	FUN12/bif-2 family
937	6	2.9	1172	2	A42587	chromopondin 2 p
938	6	2.9	1180	2	A11939	two-component hybr
939	6	2.9	1197	2	D86317	protein FISH18.21
940	6	2.9	1203	2	A59257	myosin VIIa, short
941	6	2.9	1218	2	T30889	valine-tRNA ligase
942	6	2.9	1230	2	T42735	TBP-interacting pr
943	6	2.9	1236	1	VHWVWE	structural polypro
944	6	2.9	1237	1	A34598	ecdysone-induced p
945	6	2.9	1239	1	VHWVEE	structural polypro
946	6	2.9	1240	1	VHWVEV	structural polypro
947	6	2.9	1241	1	S26373	genome polyprotein
948	6	2.9	1242	2	S72350	structural polypro
949	6	2.9	1242	2	A56605	structural polypro
950	6	2.9	1245	1	VHWVB	structural polypro
951	6	2.9	1245	1	VHWVB2	structural polypro
952	6	2.9	1245	1	VHWVB2	structural polypro
953	6	2.9	1247	1	VHWVN2	structural polypro
954	6	2.9	1248	2	C89874	autolysin (importe
955	6	2.9	1251	2	B86194	hypothetical prote
956	6	2.9	1253	1	VHWV	structural polypro
957	6	2.9	1254	1	VHWVRA	structural polypro
958	6	2.9	1254	1	VHWVVE	structural polypro
959	6	2.9	1254	1	VHWVVT	structural polypro
960	6	2.9	1254	1	J01978	structural polypro
961	6	2.9	1254	1	J01979	structural polypro
962	6	2.9	1254	2	G86379	protein F5A9.24 [I
963	6	2.9	1255	1	B44213	structural polypro
964	6	2.9	1255	1	D44213	structural polypro
965	6	2.9	1266	2	T22090	hypothetical prote
966	6	2.9	1286	2	B71413	hypothetical prote
967	6	2.9	1290	2	AE2192	two-component hybr
968	6	2.9	1291	2	T17242	hypothetical prote
969	6	2.9	1315	2	A56101	collagen alpha 1(X
970	6	2.9	1316	2	T00381	protein F25G6.9 [I
971	6	2.9	1326	2	H89134	probable membrane
972	6	2.9	1333	2	S63403	endo-1,4-beta-xyla
973	6	2.9	1347	2	T30909	hypothetical prote
974	6	2.9	1357	2	S57052	hypothetical prote
975	6	2.9	1380	2	T52284	hypothetical prote
976	6	2.9	1394	2	B34598	ecdysone-induced p
977	6	2.9	1425	2	T31153	hypothetical prote
978	6	2.9	1437	2	C75198	activator 1, repli

```

979 6 2.9 1443 2 S05979
980 6 2.9 1446 2 S73013
981 6 2.9 1508 2 B87636
982 6 2.9 1528 2 A60338
983 6 2.9 1538 2 E70874
984 6 2.9 1540 2 H87203
985 6 2.9 1545 2 T14288
986 6 2.9 1570 2 T18272
987 6 2.9 1575 2 T18345
988 6 2.9 1618 2 S21424
989 6 2.9 1643 2 T07961
990 6 2.9 1655 2 S47446
991 6 2.9 1666 2 T09072
992 6 2.9 1731 2 B88241
993 6 2.9 1731 2 A83045
994 6 2.9 1742 2 S76110
995 6 2.9 1774 2 B56101
996 6 2.9 1804 2 T14518
997 6 2.9 1805 2 A34736
998 6 2.9 1825 2 T42725
999 6 2.9 1861 2 T13845
1000 6 2.9 1885 2 T30847

```

ALIGNMENTS

```

RESULT 1
E69190
conserved hypothetical protein MTH68 - Methanobacterium thermoautotrophicum (strain Delta
C/Species: Methanobacterium thermoautotrophicum
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jan-2000
C/Accession: E69190
R/Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qu, D.; Spadafora, R.; Vicalite, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A/Reference number: A69000; MUID:96037514; PMID:9371463
A/Accession: E69190
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-228 <MTH>
A/Cross-references: GB:A6000666
A/Experimental source: strain Delta H
C/Genetics:
A/Gene: MTH68
C/Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc
F/1-34/Domain: tetratricopeptide repeat homology <TT1>
F/35-68/Domain: tetratricopeptide repeat homology <TT2>
F/69-102/Domain: tetratricopeptide repeat homology <TT3>
F/103-136/Domain: tetratricopeptide repeat homology <TT4>
F/137-170/Domain: tetratricopeptide repeat homology <TT5>
F/171-204/Domain: tetratricopeptide repeat homology <TT6>

Query Match 4.9%; Score 10; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 FEKAIEINPK 103
Db 24 FEKAIEINPK 33

```

```

J. Bacteriol. 179, 7135-7155, 1997
A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A/Reference number: A69000; MUID:96037514; PMID:9371463
A/Accession: E69190
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-379 <MTH>
A/Cross-references: GB:A6000799; GB:A6000666; NID:g2621112; PIDN:AB84589.1; PID:g26211.
A/Experimental source: strain Delta H
C/Genetics:
A/Gene: MTH68
C/Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat h
F/1-49/Domain: tetratricopeptide repeat homology <TT01>
F/50-83/Domain: tetratricopeptide repeat homology <TT02>
F/84-117/Domain: tetratricopeptide repeat homology <TT03>
F/118-151/Domain: tetratricopeptide repeat homology <TT04>
F/152-185/Domain: tetratricopeptide repeat homology <TT05>
F/186-219/Domain: tetratricopeptide repeat homology <TT06>
F/220-253/Domain: tetratricopeptide repeat homology <TT07>
F/254-287/Domain: tetratricopeptide repeat homology <TT08>
F/288-321/Domain: tetratricopeptide repeat homology <TT09>
F/322-355/Domain: tetratricopeptide repeat homology <TT10>

```

```

Query Match 4.9%; Score 10; DB 2; Length 379;
Best Local Similarity 100.0%; Pred. No. 0.056;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 FEKAIEINPK 103
Db 175 FEKAIEINPK 184

```

```

RESULT 3
T01795
hypothetical protein A.TM021B04.14 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 22-Oct-1999
C/Accession: T01795
R/Dante, M.; Wamsley, P.; Gibson, A.
Submitted to the EMBL Data Library, June 1997
A/Description: The sequence of A. thaliana TM021B04.
A/Reference number: Z14440
A/Accession: T01795
A/Status: translated from GB/EMBL/DDBT
A/Molecule type: DNA
A/Residues: 1-212 <DAN>
A/Cross-references: EMBL:A6007271; NID:g2191181; PID:g2191191; GSPDB:GN00063; ATSP:A_TM
A/Gene: ATSP:A.TM021B04.14
A/Map position: 5
A/Introns: 43/1; 69/3

Query Match 3.9%; Score 8; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AMGLVLSA 19
Db 72 AMGLVLSA 79

```

```

RESULT 4
B71441
probable myosin II heavy chain - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
A/Variety: columbiana
C/Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C/Accession: B71441
R/Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dietrich,
; Wedler, H.; Wedler, E.; Wamboldt, R.; Welteneberger, T.; Pohl, T.M.; Terry, N.; Giel
; Qu, D.; Spadafora, R.; Vicalite, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
Nature 391, 485-488, 1998
A/Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech

```


erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; And
C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MWID:98121113; PMID:9461215
A:Accession: B71441
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-550 <BEV>
A:Cross-references: GB:297343; NID:G2245073; PID:G327533; PID:G2245081
C:Genetics:
A:Map position: 4COP9-4G3845

Query Match 3.9%; Score 8; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 SAKERALI 178
DB 159 SAKERALI 166

RESULT 5
T41509
Serine/threonine-protein kinase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C:Accession: T41509
R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
Submitted to the EMBL Data Library, March 1999
A:Reference number: Z21999
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Accession: T41509
A:Molecule type: DNA
A:Residues: 1-830 <MUR>
A:Cross-references: EMBL:AL049522; PIDN:CAB40012.1; GSPDB:GN00068; SPDB:SPCC63.08C
C:Experimental source: strain 972h-; cosmid c63
C:Genetics:
A:Gene: SPDB:SPCC63.08C
A:Map position: 3
A:Introns: 34/3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 3.9%; Score 8; DB 2; Length 830;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KSALEPRL 177
DB 541 KSALEPRL 548

RESULT 6
A47213
beta-fodrin - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C:Accession: A47213
R:Chang, J.G.; Scarpa, A.; Eddy, R.L.; Byers, M.G.; Harris, A.S.; Morrow, J.S.; Watkins,
Genomics 17, 287-293, 1993
A:Title: Cloning of a portion of the chromosomal gene and cDNA for human beta-fodrin, ch
A:Reference number: A47213; MWID:94010920; PMID:8406479
A:Accession: A47213
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1252 <CHA>
A:Cross-references: GB:865762; NID:G425552; PIDN:AAB28324.1; PID:G425553
A:Note: sequence extracted from NCBI backbone (NCBIN:137919, NCBIIP:137922)
C:Superfamily: spectrin beta-G chain; alpha-actinin actin-binding domain homology; pleck
C:Keywords: actin binding
F:9-120/Domain: spectrin/dystrophin repeat homology <SP1>
F:129-234/Domain: spectrin/dystrophin repeat homology <SP2>

Query Match 3.9%; Score 8; DB 2; Length 1252;

Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 QTKAQLF 94
DB 1078 QTKAQLF 1085

RESULT 7
A44159
Spectrin beta-G chain - human
N:Alternate names: beta-spectrin general isoform, beta G-spectrin
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A44159
R:Hu, R.J.; Watanabe, M.; Bennett, V.
J. Biol. Chem. 267, 18715-18722, 1992
A:Title: Characterization of human brain cDNA encoding the general isoform of beta-spect
A:Reference number: A44159; MWID:92406787; PMID:1527002
A:Accession: A44159
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2364 <HUI>
A:Cross-references: GB:M96803; NID:G338442; PIDN:AAA60580.1; PID:G338443
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIIP:113399)
C:Genetics:
A:Gene: GDB:SPTBN1
A:Cross-references: GDB:120386; OMIM:182790
A:Map position: 2p21-2p21
C:Superfamily: spectrin beta-G chain; alpha-actinin actin-binding domain homology; pleck
C:Keywords: actin binding; cytoskeleton; duplication; heterodimer; membrane protein
F:53-271/Domain: alpha-actinin actin-binding domain homology <ACT>
F:301-412/Domain: spectrin/dystrophin repeat homology <SP1>
F:1697-1803/Domain: spectrin/dystrophin repeat homology <SP2>
F:2196-2305/Domain: pleckstrin repeat homology <PLK>

Query Match 3.9%; Score 8; DB 1; Length 2364;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 QTKAQLF 94
DB 1370 QTKAQLF 1377

RESULT 8
E69549
Conserved hypothetical protein AF2396 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: E69549
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Ariach, P.; Kaine, B.P.; Sykes, S
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MWID:98049343; PMID:9389475
A:Accession: E69549
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-78 <KLE>
A:Cross-references: GB:AE00111; GB:AE000782; NID:G2689434; PIDN:AAB91275.1; PID:G265070
C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MW0567

Query Match 3.4%; Score 7; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 EGRVAVI 144

Db 14 EGRVVAI 20

RESULT 9

D64371
hypothetical protein M05072 - Methanococcus jannaschii
C/Species: Methanococcus jannaschii
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: D64371
R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
R. Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Merrick, J.M.; Glodex, A.;
rson, J.D.; Sddow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurest, M.A.
Science 273, 1058-1073, 1996
A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A/Reference number: A64301; MUID:96337999; PMID:8688087
A/Accession: D64371
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-86 <BU>
A/Cross-references: GB:U67506; GB:L77117; NID:g1591274; PIDN:AA98572.1; PID:g1591279; T
C/Genetics:
A/Map position: FOR509862-510122
C/Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc
F/8-41/Domain: tetratricopeptide repeat homology <TT1>
F/42-75/Domain: tetratricopeptide repeat homology <TT2>

Query Match 3.4%; Score 7; DB 2; Length 86;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 KALEINP 102
|||||
Db 33 KALEINP 39

RESULT 10

AD3217
conserved hypothetical protein Atus467 [imported] - Agrobacterium tumefaciens (strain C5
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C/Accession: AD3217
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Giller, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; Mclell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kressan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; PMID:11743193
A/Accession: AD3217
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-103 <KUR>
A/Cross-references: GB:AE008687; PIDN:AA146154.1; PID:g17743924; GSPDB:GN00188
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atus467
A/Genome: plasmid

Query Match 3.4%; Score 7; DB 2; Length 103;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KSALEA 176
|||||
Db 23 KSALEA 29

RESULT 11

F70422
hypothetical protein ag_1409 - Aquifex aeolicus
C/Species: Aquifex aeolicus

C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 11-Jan-2000
C/Accession: F70422

R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

V. Nature 392, 353-358, 1998

A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Accession: F70422

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-131 <AQF>

A/Cross-references: GB:AE000739; NID:g2983813; PIDN:AA07377.1; PID:g2983821; GB:AE00065

A/Experimental source: strain VFS

C/Genetics:

A/Gene: ag_1409

C/Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc

F/43-76/Domain: tetratricopeptide repeat homology <TT1>

F/77-110/Domain: tetratricopeptide repeat homology <TT2>

Query Match 3.4%; Score 7; DB 2; Length 131;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 AYRTLAK 76
|||||
Db 79 AYRTLAK 85

RESULT 12

D59097
hypothetical protein PX01-52 - Bacillus anthracis virulence plasmid PX01

C/Species: Bacillus anthracis

C/Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 16-Feb-2001

C/Accession: D59097

R/Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler

J. Bacteriol. 181, 6509-6515, 1999

A/Title: Sequence and organization of PX01, the large Bacillus anthracis plasmid harbori

A/Reference number: A59091; MUID:99445483; PMID:10515943

A/Accession: D59097

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-149 <OKI>

A/Cross-references: GB:AF065404; NID:g4894216; PIDN:AA032356.1; PID:g4894268

A/Experimental source: strain Sterne

C/Genetics:

A/Gene: PX01-52

A/Genome: plasmid

C/Superfamily: hypothetical protein PX01-52

Query Match 3.4%; Score 7; DB 2; Length 149;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KSALEA 176
|||||
Db 35 KSALEA 41

RESULT 13

C69842

multidrug resistance protein homolog yltZ - Bacillus subtilis

C/Species: Bacillus subtilis

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

C/Accession: C69842

R/Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berrei

C.; Bron, S.; Bruggliet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Enllich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati, E.

Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Galle

iech, U.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holstappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois,

A/Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee]

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seroy, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seroy, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A.; Authors: Yoshikawa, H.F.; Zumelein, E.; Yoshikawa, H.; Darchin, A. A.; Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: C69842
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-164 <KUN>
 A:Cross-references: GB:Z99109; GB:AL009126; NID:G2633260; PIDN:CAB12959.1; PID:el183121;
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yltz

Query Match 3.4%; Score 7; DB 2; Length 164;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 LSGALTY 127
 DB 35 LSGALTY 41

RESULT 14
 S75578
 hypothetical protein slr0823 - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 C:Accession: S75578
 R:Kaneko, T.; Sato, S.; Kozami, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

8.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S75578
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-173 <KAN>
 A:Cross-references: EMBL:D90911; GB:AB001339; NID:G1653083; PIDN:BA16139.1; PID:G165322
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: chloroplast conserved hypothetical protein 167; tetratricopeptide repeat
 F:35-68/Domain: tetratricopeptide repeat homology <TR>

Query Match 3.4%; Score 7; DB 2; Length 173;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 ATELNPK 103
 DB 98 ATELNPK 104

RESULT 15
 AH2035
 hypothetical protein all1838 [imported] - *Nostoc* sp. (strain PCC 7120)
 C:Species: *Nostoc* sp.
 A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AH2035
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kunitz, T.; Sasanoto, S.; Watanabe, A.; Itiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
 A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium *Ana*
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2035
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-193 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BAH73537.1; PID:G17130928; GSPDB:GN00179

A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all1838
 C:Superfamily: *Synechocystis* hypothetical protein slr2011

Query Match 3.4%; Score 7; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ALERALI 178
 DB 128 ALERALI 134

RESULT 16
 T01040
 hypothetical protein YUP8H12R.23 - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 24-Nov-1999
 C:Accession: T01040
 R:Theologis, A.; Vysotskaya, V.S.; Osborne, B.I.; Schwartz, J.R.; Federgruel, N.A.; Kwei
 Oefner, P.; Davis, R.W.
 submitted to the EMBL Data Library, May 1998
 A:Description: *Arabidopsis thaliana* chromosome 1 YAC YUP8H12R sequence.
 A:Reference number: Z14227
 A:Accession: T01040
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-204 <THE>
 A:Cross-references: EMBL:AC002986; NID:G2494106; PID:G3152565; GSPDB:GN00059; ATSP:YUP8H
 C:Genetics:
 A:Gene: ATSP:YUP8H12R.23
 A:Map position: 1
 A:Introns: 71/3
 C:Superfamily: *Arabidopsis thaliana* hypothetical protein YUP8H12R.23

Query Match 3.4%; Score 7; DB 2; Length 204;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KSALEBA 176
 DB 122 KSALEBA 128

RESULT 17
 AE1763
 the cell-division ATP-binding protein FtsE homolog ftsE [imported] - *Listeria innocua* (s
 C:Species: *Listeria innocua*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
 C:Accession: AE1763
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fahn, H
 D.; Jones, L.M.; Kayst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Matounam, A.; Ma
 ok, C.; Schuener, T.; Simoes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of *Listeria species*.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AE1763
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-228 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CAC97877.1; PID:G16415187; GSPDB:GN00178
 A:Experimental source: strain C11P11262
 C:Genetics:
 A:Gene: ftsE
 C:Superfamily: short-chain ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 3.4%; Score 7; DB 2; Length 228;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 RVAIEN 146
|||||
Db 206 RVAIEN 212

RESULT 18

hypoetical protein SPBC2G2.09c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: T40148
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.
submitted to the EMBL Data Library, March 1998
A:Reference number: 221842
A:Accession: T40148
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-229 <MOO>
A:Cross-references: EMBL:AL022103; PIDN:CAAL789.1; GSPDB:GN00067; SPDB:SPBC2G2.09c
A:Experimental source: strain 972h-; cosmid c2G2
C:Genetics:
A:Gene: SPDB:SPBC2G2.09c
A:Map position: 2
A:Introns: 96/3; 150/1; 168/3

Query Match 3.4%; Score 7; DB 2; Length 229;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 40 IALSIL 46
|||||
Db 109 IALSIL 115

RESULT 19

hypoetical protein SA1856 (imported) - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: A89997
R:Kuroda, M.; Ohia, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Chi, L.; Oguc
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: A89997
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-229 <KUR>
A:Cross-references: GB:BA000018; PID:G13701846; PIDN:BA843138.1; GSPDB:GN00149
C:Genetics:
A:Experimental source: strain N315
A:Gene: SA1856

Query Match 3.4%; Score 7; DB 2; Length 229;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 AKOOLA 59
|||||
Db 62 AKOOLA 68

RESULT 20

unknown protein encoded by prophage CP-933X [imported] - Escherichia coli (strain O157:H
D85693
C/Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C/Accession: D85693
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <STO>
A:Cross-references: GB:AE005174; NID:G12514848; PIDN:AA656008.1; GSPDB:GN00145; UNGP:21;
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1919
C/Superfamily: phage lambda hypoetical 20.2K protein

Query Match 3.4%; Score 7; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 58 DNALSND 64
|||||
Db 85 DNALSND 91

RESULT 21

Probable tricolitoxin immunity protein [imported] - Sinorhizobium meliloti (strain 1021)
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C/Accession: G95902
R:Finan, T.M.; Weldner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: G95902
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <KUR>
A:Cross-references: GB:AL591985; PIDN:CA648887.1; PID:G15140360; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
peta, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kallman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Laure,
heault, P.; Vandebol, M.; Vorholter, F.J.; Weldner, S.; Wells, D.H.; Wong, K.; Yeh, K
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A86039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: tfixG; SMD20505
A:Genome: plasmid

Query Match 3.4%; Score 7; DB 2; Length 262;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 SALERAL 177
|||||
Db 255 SALERAL 261

RESULT 22

co-chaperone-curved DNA binding protein A - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 21-Jan-2000
C/Accession: H64647
R:Tomb, J.F.; White, O.; Kesteven, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKen
son, J.D.; Kelley, J.M.; Cotton, M.D.; Feldman, J.M.; Fujii, C.; Bowman, C.; Watney, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Frazer, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: H64647

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-288 <ROM>
A:Cross-references: GB:AE000610; GB:AE000511; NID:g2314160; PIDN:AA08066.1; PID:g231416
F/4-68/Domain: dnaJ amino-terminal homology <DNM>

Query Match
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 DEIKKSY 190
DB 18 DEIKKSY 24

RESULT 23

C71936
Probable co-chaperone with dnaJ - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 31-Jan-2000
C/Accession: C71936
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: C71936
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-288 <ARN>
A:Cross-references: GB:AE001474; GB:AE001439; NID:g4154929; PIDN:AA05985.1; PID:g415493
A:Experimental source: strain J99
C:Genetics:
A:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

Query Match
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 DEIKKSY 190
DB 18 DEIKKSY 24

RESULT 24

T19669
Hypothetical protein C3JA12.16 - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T19669
R:Milkinson, J.
Submitted to the EMBL Data Library, January 1996
A:Reference number: Z19159
A:Accession: T19669
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-292 <ML>
A:Cross-references: EMBL:Z68493; PIDN:CAA92799.1; GSPDB:GN00022; CESP:C3JA12.16
A:Experimental source: clone C3JA12
C:Genetics:
A:Gene: CESP:C3JA12.16
A:Map position: 4

Query Match
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SACQSTP 24
DB 246 SACQSTP 252

RESULT 25

G88783
protein C3JA12.16 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C/Accession: G88783
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88783
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <STO>
A:Cross-references: GB:chr_IV; PIDN:CAA92799.1; PID:g3874676; GSPDB:GN00022; CESP:C3JA12
A:Genetics:
A:Gene: C3JA12.16
A:Map position: 4

Query Match
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SACQSTP 24
DB 246 SACQSTP 252

RESULT 26

H82214
conserved hypothetical protein VC1311 [imported] - Vibrio cholerae (strain N16961 serogr
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: H82214
R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833; PMID:10952301
A:Accession: H82214
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <HEI>
A:Cross-references: GB:AE004211; GB:AE003852; NID:g9655798; PIDN:AAF4469.1; GSPDB:GN001
A:Experimental source: serogroup O1, strain N16961, biotype El Tor
C:Genetics:
A:Gene: VC1311
A:Map position: 1

Query Match
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AAKSPTK 163
DB 282 AAKSPTK 288

RESULT 27

AF3504
acetyl-CoA carboxylase (EC 6.4.1.2) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C/Accession: AF3504
R:Delvecchio, V.G.; Kaparal, V.; Redkar, R.V.; Patra, G.; Mujar, C.; Lee, T.; Ivanova,
.; Mazur, M.; Goldstein, E.; Selkov, E.; Elzer, P.H.; Haglund, S.; O'Callaghan, D.; Leites
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

```
A:Reference number: ADJ252; PMID:11756688
A:Accession: AF3504
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL53201.1; PID:gl7984075; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEJ2020
A:Map position: 1
C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain
C:Keywords: ligase

Query Match
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 FEKATL 100
Db 148 FEKATL 154

RESULT 28
ADJ233
DNA invertase [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid T1
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: ADJ233
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: ADJ233
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313 <KUR>
A:Cross-references: GB:AE008690; PIDN:AAL46282.1; PID:gl7744064; GSPDB:GN00189
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: y4Cg
A:Genome: plasmid

Query Match
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 LERALLS 179
Db 129 LERALLS 135

RESULT 29
AB0767
probable reductase rfbI rfbI [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
A>Note: this species has also been called Salmonella typhimurium
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 01-Mar-2002
C:Accession: AB0767
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AB0767
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-324 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD02456.1; PID:gl6503323; GSPDB:GN00176
C:Genetics:
A:Gene: rfbI
C:Superfamily: methane monooxygenase reductase component; cytochrome-b5 reductase homol
Query Match
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 LDALSA 63
Db 23 LDALSA 29

RESULT 30
S15303
probable CDP-6-deoxy-Delta(3,4)-glucosyl reductase (EC 1.3.1.-) - Salmonella typhimurium
N:Alternate names: hypothetical protein 7.6
C:Species: Salmonella typhimurium
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S15303
R:Jiang, X.M.; Neal, B.; Santiago, F.; Lee, S.J.; Romana, L.K.; Reeves, P.R.
Mol. Microbiol. 5, 695-713, 1991
A>Title: Structure and sequence of the rfb (O antigen) gene cluster of Salmonella serov
A:Reference number: S15296; PMID:1710759
A:Accession: S15303
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <MOL>
C:Superfamily: methane monooxygenase reductase component; cytochrome-b5 reductase homol
C:Keywords: 2Fe-2S; iron-sulfur protein; metalloprotein; oxidoreductase.
F:101-320/Domain: ferredoxin [2Fe-2S] homology <FER>
F:37,42,45,71/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 LDALSA 63
Db 23 LDALSA 29

RESULT 31
T39033
hypothetical protein SPAC6C3.09 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C:Accession: T39033
R:Devlin, K.; Churcher, C.M.; Barrett, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21750
A:Accession: T39033
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-335 <DEV>
A:Cross-references: EMBL:269731; PIDN:CA040281.1; GSPDB:GN00066; SPDB:SPAC6C3.09
A:Experimental source: strain 972h-; cosmid c6C3
C:Genetics:
A:Gene: SPDB:SPAC6C3.09
A:Map position: 1
A:Insertions: 31/3; 81/1; 100/1; 286/2
C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC6C3.09

Query Match
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GRLDOAK 54
Db 199 GRLDOAK 205
```

RESULT 32

A87300 TPR domain protein [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C/Accession: A87300

R/Merman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton

N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: A87300

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-336 <STO>

A/Cross-references: GB:AE005673; NID:g13421571; PIDN:AAK2397.1; GSPDB:GN00148

C/Genetics:

A/Gene: CC0410

Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 336; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

57 LDAAALSA 63

243 LDAAALSA 249

RESULT 33

T15727 hypothetical protein C31H1.5 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C/Accession: T15727

R/Le, T.

submitted to the EMBL Data Library, December 1995

A/Description: The sequence of C. elegans cosmid C31H1.

A/Reference number: Z18395

A/Accession: T15727

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-336 <LET>

A/Cross-references: EMBL:U02848; NID:g1125817; PID:g1125820; PIDN:AA83608.1; CESP:C31H1

C/Genetics:

A/Gene: CESP:C31H1.5

A/Intons: 31/2; 67/1; 99/3; 169/1; 235/1; 309/1

Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 336; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

95 EKAIEUN 101

64 EKAIEUN 70

RESULT 34

T08782 hypothetical protein DKFP586N1020.1 - human (fragment)

C/Species: Homo sapiens (man)

C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 03-Nov-2000

C/Accession: T08782

R/Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, March 1999

A/Reference number: Z16471

A/Accession: T08782

A/Molecule type: mRNA

A/Residues: 1-349 <NAM>

A/Cross-references: EMBL:AL050156

A/Experimental source: adult uterus; clone DKFP586N1020

C/Genetics:

A/Note: DKFP586N1020.1

C/Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat h

Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 349; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

96 KATLENP 102

152 KATLENP 158

RESULT 35

C83980 hypothetical protein BH2643 [imported] - Bacillus halodurans (strain C-125)

C/Species: Bacillus halodurans

C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C/Accession: C83980

R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir

Nucleic Acids Res. 28, 4317-4331, 2000

A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A/Reference number: A83650; MUID:20512582; PMID:11058132

A/Accession: C83980

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-362 <STO>

A/Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BAB06362.1; GSPDB:GN0

A/Experimental source: strain C-125

C/Genetics:

A/Gene: BH2643

Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 362; Pred. No. 60;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 VLSACQS 22

16 VLSACQS 22

RESULT 36

C64142 conserved hypothetical protein H10091 - Haemophilus influenzae (strain Rd KW20)

C/Species: Haemophilus influenzae

C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999

C/Accession: C64142

R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Krelavage,

; Gocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.

Science 269, 496-512, 1995

A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A/Reference number: A64000; MUID:95350630; PMID:7542800

A/Accession: C64142

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-378 <TIGR>

A/Cross-references: GB:U032694; GB:L42023; NID:g1573035; PIDN:AAC21769.1; PID:g1573042;

C/Superfamily: yhad protein

Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 378; Pred. No. 62;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

56 QLDAAIS 62

221 QLDAAIS 227

RESULT 37

AE0484 probable glycerate kinase (EC 2.7.1.31) [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C/Accession: AE0484
 R:Perkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, Nature 413, 523-527, 2001
 A>Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A/Reference number: AB0001; MUID:21470413; PMID:11586360
 A/Accession: AE0484
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-378 <KUR>
 A/Cross-references: GB:AL590842; PIDN:CAC93441.1; PID:G15981888; GSPDB:GN00175
 C/Genetics:
 A/Gene: glxK
 C/Superfamily: yhad protein
 C/Keywords: phosphotransferase

Query Match 3.4%; Score 7; DB 2; Length 378;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 QOLDPAL 61
 |||||
 Db 220 QOLDPAL 226

RESULT 38
 A95860
 Hypothetical protein [imported] - *Sinorhizobium meliloti* (strain 1021) megaplasmid pSymB
 C/Species: *Sinorhizobium meliloti*
 C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C/Accession: A95860
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A>Title: The complete sequence of the 1,683-Kb pSymB megaplasmid from the N2-fixing endo
 A/Reference number: A95842; MUID:21396508; PMID:11481431
 A/Accession: A95860
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-395 <KUR>
 A/Cross-references: GB:AL591985; PIDN:CAC48545.1; PID:G15140017; GSPDB:GN00167
 A/Experimental source: strain 1021, megaplasmid pSymB
 R:Gibbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, L.; Hyman, R.W.; Jones, T. Science 293, 668-672, 2001
 A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K. A>Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A/Reference number: A96039; MUID:21368234; PMID:11474104
 A/Contents: annotation
 C/Genetics:
 A/Gene: SMD20145
 A/Genome: plasmid

Query Match 3.4%; Score 7; DB 2; Length 395;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 AALSADR 65
 |||||
 Db 278 AALSADR 284

RESULT 39
 S46741
 glutamate 5-kinase homolog - yeast (*Saccharomyces cerevisiae*)
 N/Alternate names: protein H8179.14; protein YHR033w
 C/Species: *Saccharomyces cerevisiae*
 C/Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 19-Apr-2002
 C/Accession: S46741
 R/Du, Z.

submitted to the EMBL Data Library, May 1994
 A/Description: The sequence of *S. cerevisiae* cosmid 8179.
 A/Reference number: S46732
 A/Accession: S46741
 A/Molecule type: DNA
 A/Residues: 1-423 <DUZ>
 A/Cross-references: EMBL:U00062; NID:G488162; PIDN:AB68910.1; PID:G488175; GSPDB:GN00001
 C/Genetics:
 A/Gene: MIPS:YHR033w
 A/Cross-references: SGD:S0001075
 A/Map position: 8R
 C/Superfamily: glutamate 5-kinase

Query Match 3.4%; Score 7; DB 2; Length 423;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 SYDKLUS 195
 |||||
 Db 324 SYDKLUS 330

RESULT 40
 A12302
 seryl-tRNA synthetase [imported] - *Nostoc* sp. (strain PCC 7120)
 C/Species: *Nostoc* sp.
 A/Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C/Accession: A12302
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuch Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, DNA Res. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
 A/Reference number: AB1807; MUID:21595285; PMID:11759840
 A/Accession: A12302
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-426 <KUR>
 A/Cross-references: GB:BA000019; PIDN:BA075675.1; PID:G17133110; GSPDB:GN00179
 A/Experimental source: strain PCC 7120
 C/Genetics:
 A/Gene: seryS
 C/Superfamily: serine-tRNA ligase

Query Match 3.4%; Score 7; DB 2; Length 426;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ALERRALI 178
 |||||
 Db 177 ALERRALI 183

Search completed: July 8, 2003, 11:47:21
 Job time : 67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:34:45 ; Search time 22 Seconds
(without alignments)
384.599 Million cell updates/sec

Title: US-10-069-544-2
Perfect score: 204
Sequence: 1 MKIRVMPVMMGLVLSAC.....EIKSYDKLSDYKLSDYK 204

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	3.9	2363	1	SPEC_MOUSE
2	8	3.9	2364	1	SPEC_HUMAN
3	7	3.4	86	1	Y572_MERJA
4	7	3.4	144	1	TS22_CHICK
5	7	3.4	173	1	YCF3_SYNY3
6	7	3.4	313	1	SGT_HUMAN
7	7	3.4	314	1	SGT_RAT
8	7	3.4	330	1	RFBI_SALTY
9	7	3.4	335	1	YD59_SCHPO
10	7	3.4	378	1	GRK_HAENI
11	7	3.4	423	1	YHJ3_YEAST
12	7	3.4	430	1	SYS_SYNY3
13	7	3.4	446	1	CRB3_SCHPO
14	7	3.4	472	1	COBO_PYRKO
15	7	3.4	496	1	YFHK_ECOLI
16	7	3.4	509	1	DNAI_MYCPA
17	7	3.4	742	1	YJDA_ECOLI
18	7	3.4	794	1	SYQ_LUPUL
19	7	3.4	1076	1	CARB_ARCFU
20	6	2.9	62	1	CX51_CONTE
21	6	2.9	68	1	RL29_THEAC
22	6	2.9	88	1	DADR_MOUSE
23	6	2.9	88	1	YOR8_TTV1
24	6	2.9	89	1	RS17_LERIN
25	6	2.9	91	1	YOH1_SERWA
26	6	2.9	92	1	Y886_MERTH
27	6	2.9	93	1	YK1A_ECOLI
28	6	2.9	94	1	CH10_BACHD
29	6	2.9	94	1	CH10_BACPD
30	6	2.9	94	1	CH10_BACST
31	6	2.9	97	1	YC40_MERTU
32	6	2.9	107	1	TH10_CLOLI
33	6	2.9	108	1	RS25_LYCSE

34	6	2.9	117	1	GALA_COTTA
35	6	2.9	141	1	HBA3_XENLA
36	6	2.9	141	1	HBA3_XENTR
37	6	2.9	142	1	RS12_HALNI
38	6	2.9	144	1	YZ03_AQUAE
39	6	2.9	146	1	MOAE_RHOSH
40	6	2.9	154	1	PBP3_DROME
41	6	2.9	154	1	SSRP_ENTPA
42	6	2.9	159	1	N7BM_ARATH
43	6	2.9	161	1	YAS1_ARCFU
44	6	2.9	163	1	EMP3_MOUSE
45	6	2.9	163	1	EMP3_RAT
46	6	2.9	168	1	DADR_BOVIN
47	6	2.9	173	1	YCP3_ANASP
48	6	2.9	179	1	Y366_HAENI
49	6	2.9	184	1	ATPE_MARPO
50	6	2.9	184	1	TRIS_RABIT
51	6	2.9	186	1	TRIS_HUMAN
52	6	2.9	186	1	TRIS_RAT
53	6	2.9	197	1	LPRI_MYCTU
54	6	2.9	197	1	YDEO_BACSU
55	6	2.9	201	1	MOBA_HELPY
56	6	2.9	205	1	CER8_ECOLI
57	6	2.9	205	1	NEPI_MERTU
58	6	2.9	207	1	B2_DAUCA
59	6	2.9	208	1	COAE_DEIRA
60	6	2.9	208	1	IFZB_THEVO
61	6	2.9	209	1	MOBA_HELPY
62	6	2.9	212	1	SOEP_STROO
63	6	2.9	214	1	YLS1_HALMA
64	6	2.9	215	1	TRU9_ECOLI
65	6	2.9	220	1	HIS7_YEAST
66	6	2.9	222	1	KAD1_YEAST
67	6	2.9	225	1	YR95_CAUCR
68	6	2.9	227	1	RRE_DAUCA
69	6	2.9	228	1	LPRI_MYCTU
70	6	2.9	228	1	Y277_RICPR
71	6	2.9	230	1	YCDL_ECO57
72	6	2.9	230	1	YCDL_ECOLI
73	6	2.9	232	1	HIS7_SACKL
74	6	2.9	236	1	PUR7_PSEAE
75	6	2.9	237	1	YU05_STROM
76	6	2.9	248	1	LPRI_MYCTU
77	6	2.9	248	1	GPVJ_ANAPL
78	6	2.9	251	1	COBM_MYCTU
79	6	2.9	251	1	YCS2_MERTU
80	6	2.9	253	1	ADH_DROGU
81	6	2.9	253	1	ADH_DROMD
82	6	2.9	253	1	ADH_DROPE
83	6	2.9	253	1	ADH_DROPS
84	6	2.9	254	1	ADH_DROAM
85	6	2.9	254	1	ADH_DROSI
86	6	2.9	254	1	ADH_DROU
87	6	2.9	255	1	ADH_DROMT
88	6	2.9	255	1	ADH_DROER
89	6	2.9	255	1	ADH_DROME
90	6	2.9	255	1	ADH_DROOR
91	6	2.9	255	1	ADH_DROSI
92	6	2.9	255	1	ADH_DROTE
93	6	2.9	255	1	ADH_DROYA
94	6	2.9	256	1	ADH_DROTS
95	6	2.9	256	1	ADH_ZAPVT
96	6	2.9	263	1	YC72_HAENI
97	6	2.9	264	1	KKA2_KLEPN
98	6	2.9	266	1	ERMF_BACFR
99	6	2.9	266	1	ERMU_BACFR
100	6	2.9	268	1	INHA_MYCAV
101	6	2.9	269	1	INHA_MYCSM
102	6	2.9	269	1	INHA_MYCTU
103	6	2.9	270	1	DILA_COXBU
104	6	2.9	274	1	MK04_RAT
105	6	2.9	275	1	RRE_ARATH
106	6	2.9	276	1	KDGA_HELPY

Q9W6M9	corunrix-co
P06636	xenopus-lae
P08422	xenopus-tro
P15756	halobacteri
O66399	aquifex-aeo
O53091	rhodobacter
P54193	drosophila
P43659	enterococcu
Q9m9m9	arabidopsis
Q29211	archaeoglob
O35912	mus-musculu
O9qyws	rattus-norv
O95136	bos-tauris
O8y998	anabaena-sp
P43988	haemophilus
P06591	marichantia
P02645	oryzolaagus
P19237	homo-sapien
P13413	rattus-norv
O10785	mycobacteri
P96674	bacillus-su
P56415	helicobacte
P09882	escherichia
O57977	methanococ
P37707	daucus-caro
O9rtv73	deinococcus
O97B05	thermoplasma
O92174	helicobacte
O51917	streptomyce
P20571	haloarcula
O00738	escherichia
P06633	saccharomyc
P07170	saccharomyc
O944n5	caulobacter
P37706	daucus-caro
P71687	mycobacteri
O92d90	ricketsia
O8xau3	escherichia
P75897	escherichia
O02986	saccharomyc
O914w0	pseudomonas
O05071	streptomyce
O5ccp6	mycobacteri
P55147	anabaena-fl
O10672	mycobacteri
O58648	methanococ
O09009	drosophila
O09010	drosophila
P37473	drosophila
P07158	drosophila
P25139	drosophila
O03384	drosophila
O05114	drosophila
P28483	drosophila
P07162	drosophila
P00314	drosophila
P07163	drosophila
P07163	drosophila
P28484	drosophila
P51550	drosophila
P51552	zaprionus-t
O57423	haemophilus
P10552	klebsiella
P10337	bacteroides
Q02607	bacteroides
O07400	mycobacteri
P42829	mycobacteri
P46533	mycobacteri
O45885	coxsiella-bu
O63454	rattus-norv
O9m1x0	arabidopsis
Q92n55	helicobacte

ID	SPCO_MOUSE	STANDARD;	PRT; 2363 AA.
983	5	2.5	187 1
984	5	2.5	188 1
985	5	2.5	188 1
986	5	2.5	189 1
987	5	2.5	189 1
988	5	2.5	189 1
989	5	2.5	189 1
990	5	2.5	189 1
991	5	2.5	189 1
992	5	2.5	190 1
993	5	2.5	190 1
994	5	2.5	190 1
995	5	2.5	190 1
996	5	2.5	190 1
997	5	2.5	190 1
998	5	2.5	191 1
999	5	2.5	191 1
1000	5	2.5	191 1

ALIGNMENTS

RESULT 1

SPCO_MOUSE

ID SPCO_MOUSE STANDARD; PRT; 2363 AA.

AC 062261;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Spectrin beta chain, brain 1 (Spectrin, non-erythroid beta chain 1)

DE (beta-II spectrin) (Fodrin beta chain).

GN SPTBN1 OR SPTB2 OR SPTB2 OR SPTB2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID:10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BALB/C;

RX MEDLINE=93240985; PubMed=8479293;

RA Ma Y., Zimmer W.E., Riederer B.M., Goodman S.R.;

RT "The complete amino acid sequence for brain beta spectrin (beta fodrin): relationship to globin sequences";

RL Brain Res. Mol. Brain Res. 18:87-89(1993).

RN [2]

RP STRUCTURE BY NMR OF 2199-2304.

RX MEDLINE=94268558; PubMed=8208297;

RA Macias M.J., Musacchio A., Ponstingl H., Nilges M., Saraste M., Oecklkat H.;

RT "Structure of the pleckstrin homology domain from beta-spectrin.";

RL Mature 369:675-677(1994).

RN [3]

RP STRUCTURE BY NMR OF 2199-2304.

RX MEDLINE=97342712; PubMed=9199409;

RA Nilges M., Macias M.J., O'Donoghue S.I., Oecklkat H.;

RT "Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin.";

RL J. Mol. Biol. 269:408-422(1997).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 2199-2304.

RX MEDLINE=96030773; PubMed=7588597;

RA Hyyonen M., Macias M.J., Nilges M., Oecklkat H., Saraste M., Wilmanns M.;

RT "Structure of the binding site for inositol phosphates in a PH domain.";

RL EMBO J. 14:4676-4681(1995).

CC -1- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION, INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE CYTOSKELETON AT THE MEMBRANE.

CC -1- SUBUNIT: LIKE ERYTHROID SPECTRIN, THE SPECTRIN-LIKE PROTEINS

CC ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO TETRAMERS.

CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.

CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.

CC -1- SIMILARITY: CONTAINS 17 SPECTRIN REPEATS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch) or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M74773; AAC42040.1; --

DR PDB; 1BTN; 08-MAR-96.

DR PDB; 1MPH; 16-JUN-97.

DR MGD; MGI:98388; Sptb2.

DR InterPro; IPR001589; Actbind actinin.

DR InterPro; IPR001715; Calponin-like.

DR InterPro; IPR001849; PH.

DR InterPro; IPR002017; Spectrin.

DR InterPro; IPR001605; Spectrin_PH.

DR Pfam; PF00307; CH; 2.

DR Pfam; PF00169; PH; 1.

DR Pfam; PF00435; spectrin; 18.

DR PRINTS; PRO0683; SPECTRINPH.

DR SMART; SM00033; CH; 2.

DR SMART; SM00233; PH; 1.

DR SMART; SM00150; SPEC; 16.

DR PROSITE; PS00019; ACTININ_1; 1.

DR PROSITE; PS00020; ACTININ_2; 1.

DR PROSITE; PS00021; CH; 2.

DR PROSITE; PS50003; PH_DOMAIN; 1.

DR PROSITE; PS50003; PH_DOMAIN; 1.

KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein; 3D-structure.

KW

FT DOMAIN 1 275 ACTIN-BINDING (BY SIMILARITY).

FT DOMAIN 54 158 CH 1.

FT DOMAIN 173 275 CH 2.

FT REPEAT 276 384 SPECTRIN 1.

FT REPEAT 385 498 SPECTRIN 2.

FT REPEAT 499 608 SPECTRIN 3.

FT REPEAT 609 714 SPECTRIN 4.

FT REPEAT 715 819 SPECTRIN 5.

FT REPEAT 820 925 SPECTRIN 6.

FT REPEAT 926 1032 SPECTRIN 7.

FT REPEAT 1033 1139 SPECTRIN 8.

FT REPEAT 1140 1245 SPECTRIN 9.

FT REPEAT 1246 1350 SPECTRIN 10.

FT REPEAT 1351 1462 SPECTRIN 11.

FT REPEAT 1463 1562 SPECTRIN 12.

FT REPEAT 1563 1668 SPECTRIN 13.

FT REPEAT 1669 1775 SPECTRIN 14.

FT REPEAT 1776 1881 SPECTRIN 15.

FT REPEAT 1882 1987 SPECTRIN 16.

FT REPEAT 1988 2132 SPECTRIN 17.

FT DOMAIN 2196 2306 PH.

FT SEQUENCE 2363 AA; 274420 MW; 64C9E4BD26BEC788 CRC64;

Query Match 3.9%; Score 8; DB 1; Length 2363;

Best Local Similarity 100.0%; Pred. No. 9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 QTKARLF 94

Db 1370 QTKARLF 1377

RESULT 2

SPCO_HUMAN

ID SPCO_HUMAN STANDARD; PRT; 2364 AA.

AC Q01082; Q16057; Q60837;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Spectrin beta chain, brain 1 (Spectrin, non-erythroid beta chain 1)
 GN SPTBN1 OR SPTB2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RC TISSUE=Brain;
 RX MEDLINE=92406787; PubMed=1527002;
 RA Hu R.J., Matanabe M., Bennett V.;
 RT "Characterization of human brain cDNA encoding the general isoform of
 beta-spectrin.";
 RL J. Biol. Chem. 267:18715-18722(1992).
 RN [2]
 RP SEQUENCE OF 293-1544 FROM N.A.
 RX MEDLINE=94010920; PubMed=8406479;
 RA Chang J.G., Scarpa A., Eddy R.L., Byers M.G., Harris A.S.,
 RT Morrow J.S., Watkins P., Shows T.B., Forget B.G.;
 RT "Cloning of a portion of the chromosomal gene and cDNA for human beta-
 fodrin, the nonerythroid form of beta-spectrin.";
 RL Genomics 17:287-293(1993).
 RN [3]
 RP SEQUENCE OF 2087-2168 FROM N.A. (SHORT ISOFORM).
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=20267884; PubMed=10806113;
 RA Hayes N.V.L., Scott C., Heekens E., Ohtanian V., Maggs A.M.,
 RA Pinder J.C., Kordell E., Baines A.J.;
 RT "Identification of a novel C-terminal variant of beta1 spectrin: two
 RT isoforms of beta1 spectrin have distinct intracellular locations and
 RT activities.";
 RL J. Cell Sci. 113:2023-2034(2000).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 173-280.
 RX MEDLINE=97307247; PubMed=9164454;
 RA Carugo K.D., Bannelos S., Saraste M.;
 RT "Crystal structure of a calponin homology domain.";
 RL Nat. Struct. Biol. 4:175-179(1997).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 173-281.
 RX MEDLINE=99036861; PubMed=9817844;
 RA Bannelos S., Saraste M., Carugo K.D.;
 RT "Structural comparisons of calponin homology domains: implications
 RT for actin binding.";
 RL Structure 6:1419-1431(1998).
 CC -1- FUNCTION: FODRIN, WHICH SEEMS TO BE INVOLVED IN SECRETION,
 CC INTERACTS WITH CALMODULIN IN A CALCIUM-DEPENDENT MANNER AND IS
 CC THUS CANDIDATE FOR THE CALCIUM-DEPENDENT MOVEMENT OF THE
 CC CYTOSKELETON AT THE MEMBRANE.
 CC -1- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS ARE
 CC CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO TETRAMERS.
 CC The short form cannot bind to the axonal protein fodaxin.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
 CC short form, are produced by alternative splicing.
 CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M56803; AAA60580.1; -
 DR EMBL; S65762; AAB28324.1; -

DR EMBL; AJ005694; CA06678.1; -
 DR EMBL; AJ238723; CAB91088.1; -
 DR PDB; 1AA2; 04-FEB-98.
 DR PDB; 1KR; 26-FEB-99.
 DR Genew; HGNC:11275; SPTBN1.
 DR MIM; 162790; -
 DR InterPro; IPR001589; Actbind_actuin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001605; Spectrin_PH.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00307; CH; 2.
 DR Pfam; PF00435; spectrin; 17.
 DR PRINTS; PR00683; SPECTRINPH.
 DR SMART; SM00233; CH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00150; SPEC; 16.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS50021; CH; 2.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein;
 KW Alternative splicing; 3D-structure.
 FT DOMAIN 1 275 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 1 173 CH 1.
 FT DOMAIN 173 275 CH 2.
 FT REPEAT 276 384 SPECTRIN 1.
 FT REPEAT 385 498 SPECTRIN 2.
 FT REPEAT 499 608 SPECTRIN 3.
 FT REPEAT 609 714 SPECTRIN 4.
 FT REPEAT 715 819 SPECTRIN 5.
 FT REPEAT 820 925 SPECTRIN 6.
 FT REPEAT 926 1032 SPECTRIN 7.
 FT REPEAT 1033 1139 SPECTRIN 8.
 FT REPEAT 1140 1245 SPECTRIN 9.
 FT REPEAT 1246 1350 SPECTRIN 10.
 FT REPEAT 1351 1462 SPECTRIN 11.
 FT REPEAT 1463 1562 SPECTRIN 12.
 FT REPEAT 1563 1668 SPECTRIN 13.
 FT REPEAT 1669 1775 SPECTRIN 14.
 FT REPEAT 1776 1881 SPECTRIN 15.
 FT REPEAT 1882 1987 SPECTRIN 16.
 FT REPEAT 1988 2133 SPECTRIN 17.
 FT DOMAIN 2134 2307 PH.
 FT VARSPPLIC 2141 2168 MAETVDTSEMNGATEQRTSKESPP -> VSTRSQTYO
 FT VARSPPLIC 2169 2364 NIKNNSRRITASDQPSGL (IN SHORT ISOFORM).
 FT CONFLICT 1411 1411 MISSING (IN SHORT ISOFORM).
 FT SEQUENCE 2364 AA; 274629 MW; 0122DB3DF25872CC CRC64;
 Query Match 3.9%; Score 8; DB 1; Length 2364;
 Best Local Similarity 100.0%; Pred. No. 9.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 87 OTKQRLF 94
 DB 1370 OTKQRLF 1377
 RESULT 3
 ID Y572 METAJ STANDARD; PRT; 86 AA.
 AC 057992;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein M0572.
 GN M0572.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcales; Methanocaldococcus.
 OX NCBI_TaxID=2190;

```

RN SEQUENCE FROM N.A.
RP STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Ruhnman J.L., Nguyen D.,
RA Utechtack I.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Klenk H.-P., Fraser C.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: SOME, TO M.JANNASCHII MJ0940 AND MJ1345.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67506; AAB98572.1; -
DR TIGR: MJ0572; -
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 2.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 86 AA; 10258 MW; 4B1EFAC549045020 CRC64;

Query Match
Best Local Similarity 3.4%; Score 7; DB 1; Length 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 KALELNP 102
Db 33 KALELNP 39

RESULT 4
ID TS22_CHICK STANDARD; PRT; 144 AA.
AC 091012;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Regulatory protein TSC-22 (TGFB stimulated clone 22 homolog).
GN TSC22.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96437509; PubMed=8840165;
RA Sawada K., Agata K., Eguchi G.;
RT "Characterization of terminally differentiated cell state by
RT categorizing cDNA clones derived from chicken lens fibers."
RL Int. J. Dev. Biol. 40:531-535(1996).
CC -1- FUNCTION: MAY SERVE AS A TRANSCRIPTIONAL REPRESSOR.
CC -1- SUBUNIT: FORMS HOMODIMER OR HETERODIMER. CAN FORM AN HETERODIMER
CC WITH TIZZ2/THG-1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE TSC-22/DIP/BUN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: DA9740; BAA08574.1; -
DR HSSP: P80220; IDIP.
DR InterPro: IPR000580; TSC-22_Dip_Bun.
DR Pfam: PF01166; TSC22_1.
DR PRODOM: PD007152; TSC-22_Dip_Bun; 1.
DR PROSITE: PS01289; TSC22_1.
RT Transcription regulation; Repressor; Nuclear protein.
FT DOMAIN 77 LEUCINE-ZIPPER.
SQ SEQUENCE 144 AA; 15384 MW; 937FD74E14F2973D CRC64;

Query Match
Best Local Similarity 3.4%; Score 7; DB 1; Length 144;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 VVAIENM 147
Db 48 VVAIENM 54

RESULT 5
ID YCF3_SYNY3 STANDARD; PRT; 173 AA.
AC P74063;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Photosystem I assembly protein ycf3.
GN YCF3 OR SLR0823.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1146;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: Seems to be required for the assembly of the photosystem
CC I complex (by similarity).
CC -1- SUBCELLULAR LOCATION: CELLULAR THYLAKOID MEMBRANE-ASSOCIATED (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE YCF3 FAMILY.
CC -1- SIMILARITY: CONTAINS 3 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90911; BAA18139.1; -
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 3.
DR SMART: SM00028; TPR; 3.
DR KX Photosynthesis; Thylakoid; Membrane; Repeat; TPR repeat;
KM Complete proteome.
FT REPEAT 35 TPR 1.
FT REPEAT 72 TPR 2.
FT REPEAT 120 TPR 3.
SQ SEQUENCE 173 AA; 19878 MW; F7D90739E17324FA CRC64;

Query Match
Best Local Similarity 3.4%; Score 7; DB 1; Length 173;

```

Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 AIELNPK 103
DB 98 AIELNPK 104

RESULT 6

SGT_HUMAN
ID SGT_HUMAN STANDARD; PRT; 313 AA.
AC 043765;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Small glutamine-rich tetratricopeptide repeat-containing protein.
GN SGT.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98414513; PubMed=9740675;
RA Kordes E., Savelyeva L., Schwab M., Rommelaere J., Jauniaux J.-C.,
RA Cziepluch C.;
RT "Isolation and characterization of human SGT and identification of
RT homologues in Saccharomyces cerevisiae and Caenorhabditis elegans.";
RL Genomice 52:90-94(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20036592; PubMed=10567422;
RA Liu F.H., Wu S.J., Hu S.M., Heiao C.D., Wang C.;
RT "Specific interaction of the 70-kDa heat shock cognate protein with
RT the tetratricopeptide repeats.";
RL J. Biol. Chem. 274:34425-34432(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX Lameudin J.E., McCreedy P.M., Skowronski E., Wiswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garner J.,
RA Dangnan L., Eiler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andrieu T., Trankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carraro A.V.;
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RL [4]
RP SEQUENCE FROM N.A.
RX TISSUE=Uterus;
MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Well B., Wellenreuther R., Gassenhuber J., Glaesl S.,
RA Ansober W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Olterswalder B., Obermaier B., Tampe J., Heubner D.,
RA Mambitt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle, and Uterus;
RA Strausberg R.;
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: INTERACTS WITH NS1 FROM PARVOVIRUS H-1.
CC -1- TISSUE SPECIFICITY: UBIDUITOUS.
CC -1- SIMILARITY: CONTAINS 3 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AJ223828; CAA11565.1; -
DR EMBL; AT133129; CAB39725.1; -
DR EMBL; AC006538; AAD13117.1; -
DR EMBL; AL050156; CAB43297.2; -
DR EMBL; BC005165; AAH05165.1; -
DR EMBL; BC008885; AAH08885.1; -
DR HSSP; P53041; 1A17.
DR Genew; HGNC:10819; SGT.
DR MIM; 603419; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 3.
DR SMART; SM00026; TPR; 3.
KW Repeat; TPR repeat.
FT REPEAT 91 124 TPR 1.
FT REPEAT 125 158 TPR 2.
FT REPEAT 159 192 TPR 3.
SQ SEQUENCE 313 AA; 34063 MW; 80B3C71B41F3CB55 CRC64;

Query Match Score 7; DB 1; Length 313;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 KAIELNPK 102
DB 116 KAIELNPK 122

RESULT 7

SGT_RAT
ID SGT_RAT STANDARD; PRT; 314 AA.
AC 070593;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Small glutamine-rich tetratricopeptide repeat-containing protein.
GN SGT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98216782; PubMed=9557704;
RA Cziepluch C., Kordes E., Poltrey R., Grewenig A., Rommelaere J.,
RA Jauniaux J.-C.;
RT "Identification of a novel cellular TPR-containing protein, SGT, that
RT interacts with the nonstructural protein NS1 of parvovirus H-1.";
RL J. Virol. 72:4149-4156(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Uterus;
MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Well B., Wellenreuther R., Gassenhuber J., Glaesl S.,
RA Ansober W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Olterswalder B., Obermaier B., Tampe J., Heubner D.,
RA Mambitt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle, and Uterus;
RA Strausberg R.;
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: INTERACTS WITH NS1 FROM PARVOVIRUS H-1.
CC -1- SIMILARITY: CONTAINS 3 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Query Match 3.4%; Score 7; DB 1; Length 314;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 KALELNP 102

Db 116 KALELNP 122

RESULT 8

RFBI_SALTY

ID RFBI_SALTY STANDARD; PRT; 330 AA.

AC P26395;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE RFI protein.

GN RFI OR STM2093.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

OX NCBI_TaxID=602;

RN NCBI_TaxID=602;

RP SEQUENCE FROM N.A.

RC STRAIN=LT2;

RX MEDLINE=91260454; PubMed=1170759;

RA Jiang X.-M., Neal B., Santiago F., Lee S.J., Roman L.K., Reeves P.R.;

RT "Structure and sequence of the rfb (O antigen) gene cluster of

RL Salmonella serovar typhimurium (strain LT2).";

RT Mol. Microbiol. 5:695-713(1991).

RP SEQUENCE FROM N.A.

RC STRAIN=LT2 / SCS1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2.";

RT Nature 413:852-856(2001).

RT - (PATHWAY: O ANTIGEN SYNTHESIS IN LIPOLYSACCHARIDE BIOSYNTHESIS

CC (PROBABLY).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; X56793; CA940119.1; -

CC EMBL; AE008792; AAL20997.1; -

CC PIR; S15303; S15303.

CC HSP; P00235; 1PR.

CC SLYGene; SG10451; rfbI.

CC InterPro; IPR000564; 2P2S2 ferredoxin.

CC InterPro; IPR001834; Cyt_B5 reductase.

CC InterPro; IPR001433; Oxred FAD/NAD(P).

CC InterPro; IPR001221; Phe_hydroxylase.

CC Pfam; PF00111; fer2; 1.

CC Pfam; PF00175; NAD_binding_1.

CC Pfam; PF00970; PAD_binding_6; 1.

CC PRINTS; PR00410; PHEHYDRLASE.

CC PROSITE; PS00197; 2P2S2_FERREDOXIN; UNKNOWN 1.

CC Lipopolysaccharide biosynthesis; Complete proteome.

CC SEQUENCE 330 AA; 36582 MW; EFC1BEC17A0CC62D CRC64;

CC Query Match 3.4%; Score 7; DB 1; Length 330;

CC Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 LDALISA 63

Db 23 LDALISA 29

RESULT 9

YD59_SCHPO

ID YD59_SCHPO STANDARD; PRT; 335 AA.

AC Q10312; Q9UUL8;

DT 01-OCT-1996 (Rel. 34, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein C6C3.09 in chromosome I.

GN SPAC6C3.09.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN NCBI_TaxID=4896;

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21648401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feldwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,

RA James K., Jones I., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Wolcaert G., Aert R., Robben J., Gymnopre B.,

RA Welfjens I., Vansierle E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs W., Fritze C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leinbach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wamburt R., Purrelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Laure V., Motter S.,

RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";

RT Nature 415:871-880(2002).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL; Z69731; CAB40281.1; -

CC Hypothetical protein.

CC SEQUENCE 335 AA; 37948 MW; C1CD4S5850BAD323 CRC64;

CC Query Match 3.4%; Score 7; DB 1; Length 335;

CC Best Local Similarity 100.0%; Pred. No. 18;

CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GKLDAQ 54

Db 199 GKLDAQ 205

```

RESULT 10
GRK_HAEIN STANDARD; PRT; 378 AA.
ID GRK_HAEIN
AC P44507;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycerate kinase (EC 2.7.1.31).
GN GLKX OR H10091.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kesteven A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedlorn E., Cotton M.D.,
RA Utermack T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhman J.V., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD".
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + (R)-glycerate = ADP + 3-phospho-(R)-
CC glycerate.
CC -1- SIMILARITY: BELONGS TO THE GLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32694; AAC21769.1; -.
CC DR TIGR: H10091; -.
CC DR InterPro: IPR004381; Cons hypoth45.
CC DR InterPro: IPR003747; Glycerate_kinase.
CC DR Pfam: PF02585; DUF168; 1.
CC DR TIGRfam: TIGR00045; Cons hypoth45; 1.
CC DR Transferase: Kinase; Complete proteome.
CC KW SEQUENCE 378 AA; 39758 MW; EBC57937BD838A9 CRC64;
SQ
Query Match 3.4%; Score 7; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 OLDPALS 62
DB 221 OLDPALS 227

```

```

RESULT 11
YH03_YEAST STANDARD; PRT; 423 AA.
ID YH03_YEAST
AC P38690;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 46.9 kDa protein in SLR2-PUR2 intergenic region.
GN YHR033W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycet.
OX NCBI_TaxID=4932;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Faveille A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Lathuille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Natter M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VII."
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: BELONGS TO THE GLUTAMATE 5-KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00062; AAB68910.1; -.
CC DR PIR: S46741; S46741.
CC DR SGD: S0001075; YHR033W.
CC DR InterPro: IPR001048; Aa_kinase.
CC DR InterPro: IPR001057; GLU_Skinase.
CC DR InterPro: IPR002478; PUA.
CC DR Pfam: PF00696; aak_kinase; 1.
CC DR Pfam: PF01472; PUA; 1.
CC DR PRINTS: PRO0474; GLUK_KINASE.
CC DR SMART: SM00359; PUA; 1.
CC DR TIGRfam: TIGR01027; PROB; 1.
CC DR PROSITE: PS00902; GLUTAMATE_5_KINASE; 1.
CC DR Hypothetical protein; Transferase; Kinase; Proline biosynthesis.
CC KW SEQUENCE 423 AA; 46863 MW; C7CBF7B1773F1981 CRC64;
SQ
Query Match 3.4%; Score 7; DB 1; Length 423;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 189 SYDKLIS 195
DB 324 SYDKLIS 330

```

```

RESULT 12
SYS_SYNY3 STANDARD; PRT; 430 AA.
ID SYS_SYNY3
AC P73201;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SerY1-cRNA synthetase (EC 6.1.1.11) (Serine--cRNA ligase) (SerrS).
GN SERRS OR SLR1703.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905211;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugita M., Saito S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-serine + tRNA(Ser) = AMP + diphosphate
CC + L-seryl-tRNA(Ser).

```

```

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90904; BAA17227.1; -.
DR HSP: P34945; 1SER.
DR InterPro: IPR002106; ALCRNA_LigaseL1.
DR InterPro: IPR002314; tRNA-synt_2b.
DR InterPro: IPR002317; tRNA-synt_ser.
DR Pfam: PF00587; tRNA-synt_2b; 1.
DR Pfam: PF02403; Sety1 tRNA N; 1.
DR PRINTS: PR00981; TRNASYNTHSR.
DR TIGRFAMs: TIGR00414; sers; 1.
DR PROSITE: PS50862; AA_tRNA_LIGASE_II; 1.
DR AMINOACYL-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
KW complete proteome.
SQ SEQUENCE 430 AA; 48038 MW; 0E6A54434DB355E4 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ALERALI 178
DB 180 ALERALI 186

RESULT 13
CRB3_SCHPO STANDARD; PRT; 446 AA.
AC Q10272;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE WD-repeat protein crb3.
GN CRB3 OR SPAC13G7.08C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96070339; PubMed=9407031;
RA Saka Y., Esashi F., Matsusaka T., Mochida S., Yanagida M.;
RT "Damage and replication checkpoint control in fission yeast is ensured
RT by interactions of Crb2, a protein with BRCT motif, with Cuts and
RT Chl1."
RL Genes Dev. 11:3387-3400(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Soutouris J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsis K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moutie S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

```

```

RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel K., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wandut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano Z., Gloux S., Lelare V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas R., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cervuti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Sphakovski G.V., Usery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: TO YEAST YNL182C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB008572; BAA23358.1; -.
DR EMBL: Z69729; CA93596.1; -.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 4.
DR PRINTS: PR00320; GPROTEINRPT.
DR SMART: SM00320; WD40; 4.
DR PROSITE: PS00678; WD_REPEATS_1; 2.
DR PROSITE: PS00682; WD_REPEATS_2; 3.
DR PROSITE: PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
FT REPEAT 74 113 WD 1.
FT REPEAT 116 155 WD 2.
FT REPEAT 172 214 WD 3.
FT REPEAT 216 257 WD 4.
FT REPEAT 294 333 WD 5.
SQ SEQUENCE 446 AA; 49506 MW; 4C426880926401A1 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 SGALTYF 128
DB 107 SGALTYF 113

RESULT 14
COBQ_PYRKO STANDARD; PRT; 472 AA.
ID COBQ_PYRKO
AC 033475;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable cobyric acid synthase.
GN COBQ.
OS Pyrococcus kodakarensis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=69014;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KOD1;
RA Rahman R., Fujiwara S., Imanaka T.;
RT "Gene cloning and sequence analysis of cobyric acid synthase and
RT cobalamin (5'-phosphate) synthase from hyperthermophilic archaeon
RT Pyrococcus sp. KOD1."
RL J. Ferment. Bioeng. 83:109-112(1997).
CC -1- FUNCTION: CATALYZES AMIDATIONS AT POSITIONS B, D, E, AND G ON

```

```

CC ADENOSYLCOBYRINIC A,C-DIAMIDE, NH(2) GROUPS ARE PROVIDED BY
CC GLUTAMINE, AND ONE MOLECULE OF ATP IS HYDROGENOLYZED FOR EACH
CC AMINATION (BY SIMILARITY).
CC -1- PATHWAY: Cobalamin biosynthesis.
CC -1- SIMILARITY: BELONGS TO THE COB/CODQ FAMILY. COBQ SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: D50533; BAA21090.1; -.
CC InterPro: IPR002586; ChIA_P.
CC InterPro: IPR004459; COBQ.
CC Pfam: PF01656; ChIA_1.
CC TIGRFAMs: TIGR00313; COBQ; 1.
CC K06106; Cobalamin biosynthesis; Porphyrin biosynthesis.
SQ SEQUENCE 472 AA; 52646 MW; 7C0CE8A00E862D31 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 93 LFEKATE 99
Db 112 LFEKATE 118

RESULT 15
YFKH_ECOLI STANDARD; PRT; 496 AA.
AC P52101; P76587;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative sensor-like histidine kinase yfhk (EC 2.7.3.-).
GN YFKH OR B2556.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
CC -1- FUNCTION: PROBABLE MEMBER OF A TWO-COMPONENT REGULATORY SYSTEM
CC YFKH/YFHA. MAY ACTIVATE YFHA BY PHOSPHORYLATION.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U36841; AAA79818.1; ALT_INIT.
CC EMBL: AE000342; AAC75609.1; -.
CC Ecocore: EGI3214; yfhk.
CC InterPro: IPR003594; ATPbind_ATPase.
CC InterPro: IPR003660; HAMP.

```

```

DR InterPro: IPR004359; HIS_KIN_sig.
DR InterPro: IPR003661; His_kinA.
DR Pfam: PF00512; signal_1.
DR Pfam: PF00672; HAMP; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR SMART: SM00387; HATPase_c; 1.
DR SMART: SM00388; HSKA; 1.
DR PROSITE: PS50109; HIS_KIN; 1.
KW Hypothetical protein; Sensory transduction; Phosphorylation; Transferase; Kinase;
KW Transmembrane; Inner membrane; Cytoplasmic (POTENTIAL).
FT DOMAIN 1 34
FT TRANSMEM 5 55
FT DOMAIN 56 194
FT TRANSMEM 195 215
FT DOMAIN 216 496
FT DOMAIN 277 493
FT MOD_RES 280 280
SQ SEQUENCE 496 AA; 55699 MW; BAF13FF2C1027737 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 73 TLAKYQ 79
Db 103 TLAKYQ 109

RESULT 16
DNA_MYCPA STANDARD; PRT; 509 AA.
ID DNA_MYCPA
AC Q9L717;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chromosomal replication initiator protein dnaA.
GN DNAA.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1770;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q., Kapur V.;
RT "Genomic organization of the Mycobacterium avium subsp.
RT paratuberculosis origin of replication region."
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION
CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;
CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS
CC (DNAA BOX): 5'-TTATC(C/A)A(C/A)A-3'. DNAA BINDS TO ATP AND TO
CC ACIDIC PHOSPHOLIPIDS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNAA FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF22789; AAF3692.1; -.
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR001957; Bac_DnaA.
CC Pfam: PF00308; bac_dnaA; 1.
CC PRINTS: PR00051; DNAA.
CC SMART: SM00382; AAA; 1.
CC TIGRFAMs: TIGR00362; Dnaa; 1.
CC PROSITE: PS01008; DNAA; 1.
CC DNA replication; DNA-binding; ATP-binding.
KW NP_BIND 210 217
SQ SEQUENCE 509 AA; 56619 MW; 2472F3F644D34EC9 CRC64;

```

Query Match 3.4%; Score 7; DB 1; Length 509;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 DAALSD 64
DB 122 DAALSD 128

RESULT 17

YJDA_ECOLI STANDARD; PRT; 742 AA.
AC P16694;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein YJDA.
GN YJDA OR B4109.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burtand V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blatner F.R.;
RT "Analysis of the Escherichia coli genome VI. DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119 (1995).
RM [2]
RN SEQUENCE FROM N.A.
RC STRAIN=B;
RX MEDLINE=90170953; PubMed=2155230;
RA Chen C.-M., Ye Q.-Z., Zhu Z., Wanner B.L., Walsh C.T.;
RT "Molecular biology of carbon-phosphorus bond cleavage. Cloning and
RT sequencing of the ptn (psid) genes involved in alkyldiphosphate
RT uptake and C-P lyase activity in Escherichia coli B."
RL J. Biol. Chem. 265:4461-4471 (1990).
CC -1- FUNCTION: BELONGS TO AN OPERON INVOLVED IN ALKYLPHOSPHONATE
CC UPTAKE AND C-P LYASE. EXACT FUNCTION NOT KNOWN.
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN K12.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14003; AAA97008.1; -
DR EMBL; AE000483; AAC7070.1; -
DR EMBL; J05260; AAA24336.1; -
DR PIR; A35718; A35718.
DR Ecogene; EG11210; YJDA.
KW Alkylphosphonate uptake; Hypothetical protein; DNA-binding;
KW ATP-binding; Complete proteome.
FT NP BIND 76 83
FT DNA BIND 464 485
FT VARIANT 346 346
FT VARIANT 356 356
FT VARIANT 458 458
FT VARIANT 499 500
FT VARIANT 516 516
FT VARIANT 594 594
FT VARIANT 600 600
SQ SEQUENCE 742 AA; 84371 MW; 8A7008FC16A02EB3 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 742;
Best Local Similarity 100.0%; Pred. No. 36;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 RALISGT 181
DB 313 RALISGT 319

RESULT 18

SYQ_LUPULU STANDARD; PRT; 794 AA.
AC P52780;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
DE (GlnRS).
OS Lupinus luteus (Yellow lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_Taxid=3873;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Ventus;
RA Siatecka M., Kozek M., Barciszewski J.;
RT "Isolation and characterization of a cDNA clone encoding a plant gene
RT of aminoacyl-tRNA synthetase."
RL (in) Plant Gene Register PGR95-103.
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
CC diphosphate + L-glutaminyl-tRNA(Gln).
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91787; CAA62901.1; -
DR HSSP; P00962; IGTR.
DR InterPro: IPR004514; GlnS.
DR InterPro: IPR000924; Gln tRNA-synt_1c.
DR InterPro: IPR001412; tRNA-synt_1c.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR TIGRFAMs; TIGR00440; glnS; 1.
DR PROSITE; PS00178; AA tRNA_LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT SITE 277 287
FT SITE 505 509
FT BINDING 508 508
SQ SEQUENCE 794 AA; 90482 MW; 4C7D02C0795268B1 CRC64;

Query Match 3.4%; Score 7; DB 1; Length 794;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLDALS 62
DB 94 QLDALS 100

RESULT 19

CARB_ARCFU STANDARD; PRT; 1076 AA.
AC O28994;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR AF1274.

OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxId=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 Ketchum K.A., Dodson R.J., Gwinn M., Hickley E.K., Peterson J.D.,
 Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
 Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus S.,
 Peterson S., Reich C.I., McNeil L.R., Badger J.H., Glodek A., Zhou L.,
 Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
 Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 Mason T.M., Olsen G.V., Fraser C.M., Smith H.O., Woese C.R.,
 Venter J.C.;
 RA "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus.";
 RT Nature 390:364-370(1997).
 RL Nature 390:364-370(1997).
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 phosphate + L-glutamate + carbamoyl phosphate.
 CC -1- COFACTOR: Binds three manganese ions (By similarity).
 CC -1- PATHWAY: Arginine biosynthesis.
 CC -1- PATHWAY: Pyridine biosynthesis; first step.
 CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
 promotes the hydrolysis of glutamine to ammonia, which is used by
 the large (or ammonia) chain to synthesize carbamoyl phosphate (By
 similarity).
 CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL: AE001016; AAB89970.1; -.
 DR HSSP: P00968; IAGX.
 DR TIGR: AF1274; -.
 DR InterPro: IPR005483; CPase_L.
 DR InterPro: IPR005479; CPase_L_D2.
 DR InterPro: IPR005480; CPase_L_D3.
 DR InterPro: IPR005481; CPase_L_N.
 DR InterPro: IPR004362; MGS_1like.
 DR Pfam: PF00289; CPase_L_Chain; 2.
 DR Pfam: PF02786; CPase_L_D2; 2.
 DR Pfam: PF02787; CPase_L_D3; 1.
 DR Pfam: PF02142; MGS_1.
 DR PRINTS: PR00098; CPASE.
 DR PROSITE: PS00866; CPASE_1; 2.
 DR PROSITE: PS00867; CPASE_2; 1.
 KM Arginine biosynthesis; Pyridine biosynthesis; Ligase; Repeat;
 ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 402 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 403 555 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 556 939 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 940 1076 ALLOSTERIC DOMAIN.
 FT REPEAT 1 555
 FT REPEAT 556 1076
 FT NP_BIND 153 210 ATP (POTENTIAL).
 FT NP_BIND 303 353 ATP (POTENTIAL).
 FT METAL 285 285 MANGANESE 1 (BY SIMILARITY).
 FT METAL 299 299 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 301 301 MANGANESE 2 (BY SIMILARITY).
 FT METAL 830 830 MANGANESE 3 (BY SIMILARITY).
 FT METAL 842 842 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 1076 AA; 119469 MW; 3FC275F648E5949D CRC64;

Query Match 3.4%; Score 7; DB 1; Length 1076;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 139 GRVAIE 145
 DB 293 GRVAIE 299
 RESULT 20
 ID CX51_CONTE STANDARD; PRT; 62 AA.
 AC 09U70;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative conotoxin TXS.1 precursor.
 OS Conus textile (Cloth-of-gold cone).
 OC Neogastropoda; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
 OX NCBI_TaxId=6494;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Venom duct;
 RX MEDLINE=99452958; PubMed=10521453;
 RA Walker C.S., Steel D., Jacobsen R.B., Lirazan M.B., Cruz L.J.,
 Hooper D., Shetty R., Delacruz R.C., Nielsen J.S., Zhou L.M.,
 Bandyopadhyay P., Craig A.G., Oliveira B.M.;
 RA Walker C.S., Steel D., Jacobsen R.B., Lirazan M.B., Cruz L.J.,
 Hooper D., Shetty R., Delacruz R.C., Nielsen J.S., Zhou L.M.,
 Bandyopadhyay P., Craig A.G., Oliveira B.M.;
 RT "The T-supersfamily of conotoxins.";
 RL J. Biol. Chem. 274:30664-30671(1999).
 RN [2]
 RP ERRATUM.
 RA Walker C.S., Steel D., Jacobsen R.B., Lirazan M.B., Cruz L.J.,
 Hooper D., Shetty R., Delacruz R.C., Nielsen J.S., Zhou L.M.,
 Bandyopadhyay P., Craig A.G., Oliveira B.M.;
 RL J. Biol. Chem. 274:36030-36030(1999).
 CC -1- FUNCTION: Not known. Has only been studied at nucleotide level.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed by the venom duct.
 CC -1- SIMILARITY: BELONGS TO THE T-SUPERFAMILY OF CONOTOXINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL: AF167164; AAF03684.1; -.
 DR Toxin; Signal; Amidation.
 FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 49 POTENTIAL.
 FT PEPTIDE 50 60 PUTATIVE CONOTOXIN TXS.1.
 FT DISULFID 51 57 POTENTIAL.
 FT MOD_RES 60 60 POTENTIAL.
 FT MOD_RES 60 60 AMIDATION (G-61 PROVIDE AMIDE GROUP)
 FT MOD_RES 60 60 (POTENTIAL).
 SQ SEQUENCE 62 AA; 6657 MW; CE29803DEBD0A421 CRC64;
 Query Match 2.9%; Score 6; DB 1; Length 62;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 168 NAKSAL 173
 DB 38 NAKSAL 43
 RESULT 21
 RL29_THEAC STANDARD; PRT; 68 AA.
 AC P58086;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 50S ribosomal protein L29P.
 GN RPL29P OR TAI264.1
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 CC Thermoplasmataceae; Thermoplasma.
 NCBI_TaxID=2303;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RT Wewers H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;
 "The genome sequence of the thermophilic scavenger Thermoplasma
 acidophilum.";
 RL Nature 407:508-513(2000).
 CC -1- SIMILARITY: BELONGS TO THE L29P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL445067; -; NOT ANNOTATED CDS.
 DR InterPro; IPR001854; Ribosomal_L29.
 DR Pfam; PF00831; Ribosomal_L29; 1.
 DR TIGRPFAM; TIGR00012; L29; 1.
 DR PROSITE; PS00579; RIBOSOMAL_L29; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 68 AA; 7982 MW; F81A8B28320A7219 CRC64;
 QY Query Match 2.9%; Score 6; DB 1; Length 68;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 174 ERALIS 179
 |||||
 33 ERAALS 38
 DB
 RESULT 22
 DADR_MOUSE STANDARD; PRT; 88 AA.
 ID DADR_MOUSE
 AC Q61616;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE D(1A) dopamine receptor (Fragment).
 GN DRD1 OR DRD1A OR GPCR15.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Testis;
 RX MEDLINE=94116980; PubMed=8288218;
 RA Walkie T.M., Chen Y., Gilbert D.J., Moore K.J., Yu L., Simon M.I.,
 RT Copeland N.G., Jenkins N.A.;
 "Identification, chromosomal location, and genome organization of
 RT mammalian G-protein-coupled receptors.";
 RL Genomics 18:175-184(1993).
 CC -1- FUNCTION: THIS IS ONE OF THE FIVE TYPES (D1 TO D5) OF RECEPTORS
 CC FOR DOPAMINE. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G
 CC PROTEINS WHICH ACTIVATE ADENYLYL CYCLASE.
 CC -1- SUBUNIT: Interacts with calycon (by similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L20336; AAA16848.1; -
 DR MCD; MG1:99578; Drida.
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; PARTIAL.
 DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Multigene family; Lipoprotein; Palmitate.
 FT NON_TER 1 26 5 (POTENTIAL).
 FT TRANSMEM 1 26 5 (POTENTIAL).
 FT DOMAIN 27 80 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 81 >88 6 (POTENTIAL).
 FT NON_TER 88 88
 SQ SEQUENCE 88 AA; 9811 MW; 0C470D867BED0949 CRC64;
 QY Query Match 2.9%; Score 6; DB 1; Length 88;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 171 SALERA 176
 |||||
 37 SALERA 42
 DB
 RESULT 23
 YOR8_TTV1 STANDARD; PRT; 88 AA.
 ID YOR8_TTV1
 AC P19283;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE Hypothetical 10.0 kDa protein.
 OS Thermoproteus tenax virus 1 (strain KRA1) (TTV1).
 CC Viruses; dsDNA viruses, no RNA stage; Lipochiriviridae;
 CC Lipochiriviruses.
 OC NCBI_TaxID=10480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Neumann H.;
 RL Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X14855; CAA32977.1; -
 DR Hypothetical protein.
 KW SEQUENCE 88 AA; 10043 MW; CABACB34D4FADF7F CRC64;
 SQ
 QY Query Match 2.9%; Score 6; DB 1; Length 88;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 17 LSACOS 22
 |||||
 68 LSACOS 73
 DB
 RESULT 24
 RS17_LEPIN

```

ID RS17 LEPIIN STANDARD; PRT; 89 AA.
AC O9XD37;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 30S ribosomal protein S17.
GN RPSO.
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sevovar Lai;
RA MEDLINE=20088835; PubMed=10620683;
RT Zuercher R.L., Hartkeerl R.A., van de Kemp H., Bal A.E.;
RT "Characterization of the Leptospira interrogans S10-spc-alpha
operon."
RL FEMS Microbiol. Lett. 182:303-308(2000).
CC -1- FUNCTION: PROTEIN S17 BINDS SPECIFICALLY TO THE 5' END OF 16S
CC RIBOSOMAL RNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S17P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF15283; AAD40592.1; -
DR HSSP; P23828; IRP.
DR InterPro: IPR000266; Ribosomal_S17.
DR Pfam: PF00366; Ribosomal_S17_1.
DR PRINTS; PR00973; RIBOSOMAL_S17.
DR PRODOM; PD001295; Ribosomal_S17; 1.
DR PROSITE; PS00056; RIBOSOMAL_S17; 1.
KW Ribosomal protein; rRNA-binding.
SQ SEQUENCE 89 AA; 10377 MW; 3376EF12PA931AB6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 YEGRVV 142
DB 14 YEGRVV 19

RESULT 25
YOHL_SERMA STANDARD; PRT; 91 AA.
ID YOHL_SERMA STANDARD; PRT; 91 AA.
AC P55757;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 10.1 kDa protein in bioA 5' region.
OS Seratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Seratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sr41;
RA Sakurai N., Imai Y., Akatsuka H., Kawai E., Komatsubara S., Tosa T.;
RA Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP IDENTIFICATION.
RA Rudd K.E.;
RC Unpublished observations (MAR-1996).
CC -1- SIMILARITY: TO E.COLI YAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17468; -; NOT ANNOTATED_CDS.
DR InterPro: IPR003735; DUF156.
DR Pfam; PF02583; DUF156; 1.
KW Hypothetical protein.
SQ SEQUENCE 91 AA; 10102 MW; 4E442754797D69F7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 91;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 ALERAL 177
DB 24 ALERAL 29

RESULT 26
Y886_METHH STANDARD; PRT; 92 AA.
ID Y886_METHH STANDARD; PRT; 92 AA.
AC O26972;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MTH886.
GN MTH886.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Carnuso A., Bush D., Saefer H., Patwell J., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Petrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics."
RT J. Bacteriol. 179:7135-7155(1997).
CC -1- SIMILARITY: TO M.JANNASCHII M0702.1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000864; AAB85384.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 92 AA; 10498 MW; 01214D51FD7CF2B3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 KPSRAI 135
DB 56 KPSRAI 61

RESULT 27

```

YKIA_ECOLI
ID YKIA_ECOLI STANDARD; PRT; 93 AA.
AC P75704;
RT 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN YKIA OR B0392.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
NX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G., Bloch C.A., Perma N.T., Burland V.,
Riley M., Collado-Vides J., Glaeser J.D., Rode C.K., Mayhew G.F.,
RA Gregor U., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
CC -1- SIMILARITY: TO THE C-TERMINAL OF E. COLI YBBD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A600145; AAC73495.1; -.
DR EcoGene; EGI4282; YKIA.
SQ SEQUENCE 93 AA; 10618 MW; 8CBE62D536AD7B40 CRC64;
KW Hypothetical protein; Complete proteome.
Query Match 2.9%; Score 6; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 NNPOLA 34
DB 38 NNPOLA 43
ID CH10_BACHD STANDARD; PRT; 94 AA.
AC 050304;
RT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN 10 kDa chaperonin (protein Cpn10) (groES protein).
OS GROES OR GROES OR MOBP OR BH0561.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20515582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
RN [2]
RP SEQUENCE OF 7-94 FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=97141316; PubMed=8987660;
RA Xu Y., Kobayashi T., Kudo T.;

RT "Molecular cloning and nucleotide sequence of the groEL gene from the
RT alkaliphilic Bacillus sp. strain C-125 and reactivation of thermally
RT inactivated alpha-glucosidase by recombinant GroEL.";
RL Biosci. Biotechnol. Biochem. 60:1633-1636(1996).
CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
CC THE ATPASE ACTIVITY OF THE LATTER.
CC -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC (BY SIMILARITY).
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001508; BAB04280.1; -.
DR EMBL: D55630; BAA09493.1; -.
DR HSSP; P05380; 1AON.
DR InterPro: IPR001476; Chaperin_Cpn10.
DR Pfam: PF00166; cpn10; 1.
DR PRINTS; PR00297; CHAPERONIN10.
DR ProDom; PD000566; Chaperin_Cpn10; 1.
DR PROSITE; PS00681; CHAPERONINS CPN10; 1.
KW Chaperone; Heat shock; Complete proteome.
SQ SEQUENCE 94 AA; 10220 MW; 5BC6D35ADD0851A6 CRC64;
Query Match 2.9%; Score 6; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 138 EGRVVA 143
DB 36 EGRVVA 41
ID CH10_BACP3 STANDARD; PRT; 94 AA.
AC P26210;
RT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN 10 kDa chaperonin (protein Cpn10) (groES protein) (Heat shock 12 kDa
GN protein).
OS GROES OR GROES.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NX NCBI_TaxID=70306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91354309; PubMed=1679330;
RA Tamada H., Ohta T., Hamamoto T., Otawara-Hamamoto Y., Yanagi M.,
RA Hiraoka H., Hirata H., Kagawa Y.;
RT "Gene structure of heat shock proteins 61kDa and 12kDa (thermophilic
RT chaperonins) of thermophilic bacterium PS3.";
RL Biochem. Biophys. Res. Commun. 179:565-571(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92121292; PubMed=8096382;
RA Ohta T., Honda K., Saito K., Hayashi H., Tano H., Hamamoto T.,
RA Kagawa Y.;
RT "Heat shock promoter of thermophilic chaperonin operon.";
RL Biochem. Biophys. Res. Commun. 191:550-557(1993).
CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
CC THE ATPASE ACTIVITY OF THE LATTER.
CC -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; S57424; AAB25914.1; -
DR PIR; J01194; J01194.
DR PIR; J01479; J01479.
DR HSSP; P05380; 1AON.
DR InterPro; IPR001476; Chaprinin_Cpn10.
DR Pfam; PF00166; cpn10; 1.
DR PRINTS; PR00297; CHAPERONIN10.
DR ProDom; PD000566; Chaprinin_Cpn10; 1.
DR PROSITE; PS00681; CHAPERONIN_CPN10; 1.
KM Chaperone.
SQ SEQUENCE 94 AA; 10340 MW; 2A9A4954AE0B457C CRC64;

Query Match 2.9%; Score 6; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 EGRVVA 143
DB 36 EGRVVA 41

RESULT 30
ID CH10_BACST STANDARD; PRT; 94 AA.
AC 007200;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 10 kDa chaperonin (Protein Cpn10) (GROES protein).
GN GROES OR GROES.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NUB36;
RX MEDLINE=93224474; PubMed=8096841;
RA Schoen U., Schumann W.;
RT "Molecular cloning, sequencing, and transcriptional analysis of the
RT groEL operon from Bacillus stearothermophilus";
RL J. Bacteriol. 175:2465-2469 (1993).
CC -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSSES
CC THE ATPASE ACTIVITY OF THE LATTER.
CC -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC (BY SIMILARITY).
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L10137; AAA22751.2; -
DR HSSP; P05380; 1AON.
DR InterPro; IPR001476; Chaprinin_Cpn10.
DR Pfam; PF00166; cpn10; 1.
DR PRINTS; PR00297; CHAPERONIN10.
DR ProDom; PD000566; Chaprinin_Cpn10; 1.
DR PROSITE; PS00681; CHAPERONIN_CPN10; 1.
KM Chaperone; Heat shock.
SQ SEQUENCE 94 AA; 10241 MW; 88732A0C42114821 CRC64;

Query Match 2.9%; Score 6; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 EGRVVA 143
DB 36 EGRVVA 41

RESULT 31
ID YC40_METUA STANDARD; PRT; 97 AA.
AC 058637;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M01240.
GN M01240.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Ullrichback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Huret M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073 (1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U67564; AAB99245.1; -
DR TIGR; M01240; -
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 97 AA; 11401 MW; 2F4F613EA65576D7 CRC64;

Query Match 2.9%; Score 6; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 EIKKSY 190
DB 2 EIKKSY 7

RESULT 32
ID TH10_CLOLI STANDARD; PRT; 107 AA.
AC P21609;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thiodoxin (TRX).
GN TRX.
OS Clostridium litoreale (Bacterium W6).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.

CC NCBI_TaxID=1557;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=DSM 5388;
CC MEDLINE=96036738; PubMed=85296640;
CC Kreimer S., Andreesen J.R.;
CC "Glycine reductase of Clostridium litoreale. Cloning, sequencing, and
CC molecular analysis of the gdrA operon that contains two in-frame TGA
CC codons for selenium incorporation.";
CC Eur. J. Biochem. 234:192-199(1995).
CC [2]
CC PRELIMINARY SEQUENCE OF 1-45.
CC MEDLINE=91139594; PubMed=1995593;
CC Meyer M., Dietrichs D., Schmidt B., Andreesen J.R.;
CC "Thioredoxin elicits a new dihydroliipoamide dehydrogenase activity by
CC interaction with the electron-transferring flavoprotein in
CC Clostridium litoreale and Escherichia acidiphilum.";
CC J. Bacteriol. 173:1509-1513(1991).
CC -1- FUNCTION: Participates in various redox reactions through the
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U24268; AAC43576.1; -
CC HSPSP; P10599; 117R.
CC InterPro: IPR000063; Thiorod.
CC Pfam: PF00085; thiorod. 1.
CC PROSITE: PS00194; THIOREDOXIN; 1.
CC Redox-active center; Electron transport.
CC DISULFID 29 32 REDOX-ACTIVE (BY SIMILARITY).
CC SEQUENCE 107 AA; 11874 MW; 7F1117FDBF3FAF9 CRC64;
CC
CC Query Match 2.9%; Score 6; DB 1; Length 107;
CC Best Local Similarity 100.0%; Pred. No. 75;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 137 YEGRVV 142
CC |||||
CC 46 YEGRVV 51
CC
CC RESULT 33
CC RS25_LYCES
CC ID RS25_LYCES STANDARD; PRT; 108 AA.
CC AC P46301;
CC DT 01-NOV-1995 (Rel. 32; Created)
CC DT 01-NOV-1995 (Rel. 32; Last sequence update)
CC DT 15-DEC-1998 (Rel. 37; Last annotation update)
CC DE 40S ribosomal protein S25.
CC RPS25.
CC OS Lycopersicon esculentum (Tomato).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
CC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
CC NCBI_TaxID=4081;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=cv. Rutgers; TISSUE=Leaf;
CC Werner R., Guitton M.C., Muehlbach H.P.;
CC "A tomato cDNA encodes a protein homologous to the eukaryotic
CC ribosomal protein S25.";
CC Physiol. Veg. 33:373-377(1995).
CC -1- SIMILARITY: BELONGS TO THE S25E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X76714; CAAS4132.1; -
CC InterPro: IPR004977; Ribosomal_S25.
CC Pfam: PF03297; Ribosomal_S25; 1.
CC Ribosomal protein.
CC SEQUENCE 108 AA; 11914 MW; BD945BAB8A9B6849 CRC64;
CC
CC Query Match 2.9%; Score 6; DB 1; Length 108;
CC Best Local Similarity 100.0%; Pred. No. 76;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 190 YDKLIS 195
CC |||||
CC 48 YDKLIS 53
CC
CC RESULT 34
CC GALA_COTUA
CC ID GALA_COTUA STANDARD; PRT; 117 AA.
CC AC Q9W6M9; Q9W6M8;
CC DT 15-JUN-2002 (Rel. 41; Created)
CC DT 15-JUN-2002 (Rel. 41; Last sequence update)
CC DT 15-JUN-2002 (Rel. 41; Last annotation update)
CC DE Galanin precursor [Contains: Galanin; Galanin message-associated
CC peptide (GMAP)].
CC GAL.
CC OS Cutunix coturnix japonica (Japanese quail).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianine;
CC Coturnix.
CC NCBI_TaxID=93934;
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Brain;
CC MEDLINE=20557635; PubMed=11108248;
CC Sakamoto H., Ubuoka T., Kohchi C., Li D., Ukena K., Tautsui K.;
CC "Existence of galanin in lumbosacral sympathetic ganglionic neurons
CC that project to the quail uterine oviduct.";
CC Endocrinology 141:4402-4412(2000).
CC -1- FUNCTION: CONTRACTS SMOOTH MUSCLE OF THE GASTROINTESTINAL AND
CC GENITOURINARY TRACT. REGULATES GROWTH HORMONE RELEASE, MODULATES
CC INSULIN RELEASE, AND MAY BE INVOLVED IN THE CONTROL OF ADRENAL
CC SECRETION (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms, 1 (shown here) and 2; are
CC produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE GALANIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF141936; AAD37349.1; -
CC EMBL: AF141935; AAD37348.1; -
CC InterPro: IPR001600; Galanin.
CC Pfam: PF01296; Galanin; 1.
CC PRINTS: PR00273; GALANIN.
CC ProDom: PD005962; GALANIN.
CC PROSITE: PS00861; GALANIN; 1.
CC Hormone; Neuropeptide; Cleavage on pair of basic residues; Amidation;
CC Signal; Alternative splicing.
CC SIGNAL 1 19 POTENTIAL.
CC PROPEP 20 30 BY SIMILARITY.

FT PEPTIDE 33 61 GALANIN.
 FT PEPTIDE 65 117 GALANIN MESSAGE-ASSOCIATED PEPTIDE.
 FT MOD_RES 61 61 AMIDATION (G-62 PROVIDE AMIDE GROUP) (BY
 FT VARSPLIC 46 46 H -> RRIIDHLIKEMPIARGEARGAY (IN ISOFORM
 FT SEQUENCE 117 AA; 12911 MW; 8A4795C460CD4FEF CRC64;
 2.9%; Score 6; DB 1; Length 117;
 Best local Similarity 100.0%; Pred. No. 81;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 GLVLSA 19
 |||||
 Db 23 GLVLSA 28

RESULT 35
 ID HB33 XENLA STANDARD; PRT; 141 AA.
 AC P06636;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hemoglobin alpha-3 chain (Alpha-T3).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 NC NCB1_TaxID=8355;
 RX MEDLINE=85297748; PubMed=2993998;
 RN SEQUENCE FROM N.A.
 RA Banville D., Williams J.G.;
 RT "The pattern of expression of the Xenopus laevis tadpole alpha-globin
 gene and the amino acid sequence of the three major tadpole alpha-
 globin polypeptides.";
 RL Nucleic Acids Res. 13:5407-5421 (1985).
 CC -1- FUNCTION: THIS IS A LARVAL (TADPOLE) ALPHA-GLOBIN.
 CC -1- SUBUNIT: Heterotrimer of two alpha chains and two beta chains.
 CC -1- TISSUE SPECIFICITY: Red blood cells.
 CC -1- POLYMORPHISM: ALPHA T3 MAY BE AN ALLELE OF ALPHA T4.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X02796; CAA28564.1; -.
 DR PIR: A24338; A24338.
 DR HSSP: P01922; 1ABW.
 DR InterPro: IPR002338; Alpha haem.
 DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; globin.1.
 DR PRINTS: PR00612; ALPHAHAEM.
 DR PROSITE: PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 58 58 IRON (HEME DISTAL LIGAND).
 FT METAL 87 87 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 141 AA; 15303 MW; F0F16943687C0A7 CRC64;

Qy 57 LDPAALS 62
 |||||
 Db 76 LDPAALS 81

RESULT 36
 ID HB33 XENTR STANDARD; PRT; 141 AA.
 AC P08422;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hemoglobin alpha-3 chain (larval).
 OS Xenopus tropicalis (western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Silurana.
 NC NCB1_TaxID=8364;
 RX MEDLINE=8157727; PubMed=3347501;
 RA Knoechel W., Pfanne K., Beck J., Meyerhof W.;
 RT "Nucleotide sequence of a larval alpha globin gene from Xenopus
 tropicalis.";
 RL Nucleic Acids Res. 16:1625-1625 (1988).
 CC -1- FUNCTION: THIS IS A LARVAL (TADPOLE) ALPHA-GLOBIN.
 CC -1- SUBUNIT: Heterotrimer of two alpha chains and two beta chains.
 CC -1- TISSUE SPECIFICITY: Red blood cells.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X06664; CAA29864.1; -.
 DR PIR: A28538; A28538.
 DR HSSP: P01966; 1FSX.
 DR InterPro: IPR002338; Alpha haem.
 DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; globin.1.
 DR PRINTS: PR00612; ALPHAHAEM.
 DR PROSITE: PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Erythrocyte.
 FT INIT MET 0 0
 FT METAL 58 58 IRON (HEME DISTAL LIGAND).
 FT METAL 87 87 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 141 AA; 15355 MW; 01C7C9F83973FCDA CRC64;

Query Match 2.9%; Score 6; DB 1; Length 141;
 Best local Similarity 100.0%; Pred. No. 95;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 LDPAALS 62
 |||||
 Db 76 LDPAALS 81

RESULT 37
 ID RS12 HALN1 STANDARD; PRT; 142 AA.
 AC P15756; Q9H82;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 30S ribosomal protein S12P (HmaS12).
 GN RPS12P OR VNC2658G.
 OS Halobacterium sp. (strain NRC-1), and
 OS Halobacterium halobium.
 OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 NC NCB1_TaxID=64091, 2242;
 [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RA NG W.V., Kennedy S.P., Mahaitas G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Bailga N.S., Thorsson V., Shrogha J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
 RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Meddocks D.G., Jablonki P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenberger T.A., Peck R.F., Fohlischoder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
 RT Genome sequence of Halobacterium species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=H.halobium; STRAIN=R1;
 RX MEDLINE=89198633; PubMed=2495365;
 RA Leffers H., Gropp F., Loetsch F., Zillig W., Garrett R.A.;
 RT "Sequence, organization, transcription and evolution of RNA
 RT polymerase subunit genes from the archaeobacterial extreme halophiles
 RT Halobacterium halobium and Halococcus morrhuae.";
 RL J. Mol. Biol. 206:1-17(1989).
 CC -1- FUNCTION: PROTEIN S12 IS INVOLVED IN THE TRANSLATION INITIATION
 CC STEP.
 CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE005138; AAG20689.1; -
 DR EMBL; AE005138; AAG20689.1; -
 DR PIR; S03581; S03581.
 DR InterPro; IPR000230; Ribosomal_S12.
 DR Pfam; PF00164; Ribosomal_S12; 1.
 DR ProDom; PD000576; Ribosomal_S12; 1.
 DR TIGRfam; TIGR00982; S23_S12_E_A; 1.
 DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 142 AA; 15489 MW; 8F426668942DBCE4 CRC64;
 Query Match 2.9%; Score 6; DB 1; Length 142;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 118 MGDLG 123
 Db 111 MGDLG 116
 RESULT 38
 ID YZ03 AQUAE STANDARD; PRT; 144 AA.
 AC 066359;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AA03.
 GN AA03.
 OS Aquifex aeolicus.
 OC Plasmid ecel.
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
 OC Aquifex.
 OX NCB1_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VFS;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,

RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.";
 RL Nature 392:353-358(1998).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE00667; AAC07951.1; -
 DR EMBL; AE00667; AAC07951.1; -
 DR KW Hypothetical protein; Plasmid; Complete proteome.
 SQ SEQUENCE 144 AA; 16600 MW; A58D8DC538E9F1 CRC64;
 Query Match 2.9%; Score 6; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 95 EXAIEL 100
 Db 85 EXAIEL 90
 RESULT 39
 ID MOAE RHOSH STANDARD; PRT; 146 AA.
 AC 053091;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Molybdopterin converting factor subunit 2 (MPT synthase subunit 2)
 DE Molybdopterin synthase subunit 2 (Molybdenum cofactor biosynthesis
 DE protein E) (Molybdopterin converting factor large subunit).
 GN MOAE.
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
 OC Rhodospirillum.
 OX NCB1_TaxID=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
 RX MEDLINE=96165256; PubMed=8576035;
 RA Dryden S.C., Downham W.;
 RT "Isolation and expression of the Rhodospirillum rubrum (pgsA)
 RT encoding phosphatidylglycerophosphate synthase.";
 RL J. Bacteriol. 178:1030-1038(1996).
 CC -1- FUNCTION: Converts molybdopterin precursor Z into molybdopterin.
 CC This requires the incorporation of two sulfur atoms into precursor
 CC Z to generate a dithiolene group (by similarity).
 CC -1- PATHWAY: Molybdenum cofactor biosynthesis.
 CC -1- SUBUNIT: Heterodimer of 2 moae subunits and 2 moae subunits (by
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE MOAE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U29587; AAC44004.1; -
 DR HSSP; P30749; IFM0.
 DR InterPro; IPR003448; Mb_biosynth_MOAE.
 DR Pfam; PF02391; Moae; 1.
 KW Molybdenum cofactor biosynthesis.
 SQ SEQUENCE 146 AA; 15779 MW; B2DAB3FB8DD22B0 CRC64;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using bw model

Run on: July 8, 2003, 11:42:00 ; Search time 81 Seconds

(without alignments)
518.933 Million cell updates/sec

Title: US-10-069-544-2

Perfect score: 204
Sequence: 1 MKIRVKMPVMVMMGLVLSAC.....EIKKSYDKLLSDYKLSDPK 204

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL 21:*

1: sp_archaea:*\n2: sp_bacteria:*\n3: sp_fungi:*\n4: sp_human:*\n5: sp_invertebrate:*\n6: sp_mammal:*\n7: sp_mhc:*\n8: sp_organelle:*\n9: sp_phage:*\n10: sp_plant:*\n11: sp_rodent:*\n12: sp_virus:*\n13: sp Vertebrate:*\n14: sp_unclassified:*\n15: sp_virus:*\n16: sp_bacteriap:*\n17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	5.4	305	11	Q9DCV4
2	11	5.4	305	11	Q9CWS6
3	11	5.4	305	11	Q8RI66
4	11	5.4	314	4	Q9Y398
5	11	5.4	314	4	Q96DB5
6	10	4.9	379	17	Q26186
7	8	3.9	212	10	O04646
8	8	3.9	397	13	Q9DDRL
9	8	3.9	515	10	Q9M0M6
10	8	3.9	550	10	O23565
11	8	3.9	830	3	Q9Y774
12	8	3.9	2154	11	Q9QWJ7
13	7	3.4	758	17	O30275
14	7	3.4	107	16	Q8UUK9
15	7	3.4	120	17	Q8TYX6
16	7	3.4	121	17	Q97610

17	7	3.4	131	16	O67408	O67408 aquifex aeo
18	7	3.4	149	2	Q9X322	Q9X322 bacillus an
19	7	3.4	156	2	O44475	O44475 azotobacter
20	7	3.4	164	16	P70952	P70952 bacillus su
21	7	3.4	165	17	Q97AS8	Q97AS8 thermoplas
22	7	3.4	176	10	Q9M5Q0	Q9M5Q0 oryza sativ
23	7	3.4	183	16	Q98B31	Q98B31 rhizobium 1
24	7	3.4	190	2	Q9R484	Q9R484 agrobacteri
25	7	3.4	193	16	O9A1R0	O9A1R0 anabaena sp
26	7	3.4	204	10	O64536	O64536 arabidopsis
27	7	3.4	215	2	Q9JEN0	Q9JEN0 neisseria m
28	7	3.4	215	2	Q9KX57	Q9KX57 mycoplasma
29	7	3.4	224	2	O9ZHL5	O9ZHL5 desulfoloma
30	7	3.4	228	16	Q927Y6	Q927Y6 listeria in
31	7	3.4	229	3	O43008	O43008 schizosacch
32	7	3.4	229	16	Q99SK1	Q99SK1 streptomyce
33	7	3.4	241	16	O8Y225	O8Y225 raietonia s
34	7	3.4	249	16	O8X4U5	O8X4U5 escherichia
35	7	3.4	262	16	Q92W58	Q92W58 rhizobium m
36	7	3.4	288	16	O25668	O25668 helicobacte
37	7	3.4	288	16	Q92M23	Q92M23 helicobacte
38	7	3.4	291	16	Q9KX2	Q9KX2 streptomyce
39	7	3.4	292	5	O18362	O18362 caenorhabdi
40	7	3.4	296	16	Q9KSE6	Q9KSE6 vibrio chol
41	7	3.4	296	2	O87552	O87552 bacillus ps
42	7	3.4	299	16	O8XK94	O8XK94 raietonia s
43	7	3.4	301	16	O8Y5E8	O8Y5E8 brucella me
44	7	3.4	313	2	Q9R705	Q9R705 agrobacteri
45	7	3.4	313	2	Q9R6N7	Q9R6N7 agrobacteri
46	7	3.4	315	11	Q99L52	Q99L52 mus musculu
47	7	3.4	324	16	O82513	O82513
48	7	3.4	329	4	O9B129	O9B129 homo sapien
49	7	3.4	334	10	Q9XEN7	Q9XEN7 titicum ae
50	7	3.4	336	5	O18339	O18339 caenorhabdi
51	7	3.4	336	16	Q9AB24	Q9AB24 caulobacter
52	7	3.4	348	16	O8U689	O8U689 agrobacteri
53	7	3.4	360	5	O9W1X6	O9W1X6 drosophila
54	7	3.4	362	16	Q9K9K3	Q9K9K3 bacillus ha
55	7	3.4	374	2	O9F5B2	O9F5B2 agrobacteri
56	7	3.4	378	16	O8ZM40	O8ZM40 yersinia pe
57	7	3.4	395	16	Q92X17	Q92X17 rhizobium m
58	7	3.4	401	2	Q9RBG0	Q9RBG0 alcaligenes
59	7	3.4	426	16	O8Y059	O8Y059 anabaena sp
60	7	3.4	432	16	O8Y735	O8Y735 anabaena sp
61	7	3.4	452	17	O97V68	O97V68 sulfolobus
62	7	3.4	458	16	P72963	P72963 synechocyst
63	7	3.4	461	16	O8Z417	O8Z417 salmoneila
64	7	3.4	479	3	O14428	O14428 neurospora
65	7	3.4	479	3	P94535	P94535 bacillus su
66	7	3.4	480	16	O8ZM30	O8ZM30 salmoneila
67	7	3.4	494	5	O9W3E0	O9W3E0 caenorhabdi
68	7	3.4	496	16	O8X447	O8X447 escherichia
69	7	3.4	498	2	Q9KWA9	Q9KWA9 agrobacteri
70	7	3.4	530	10	Q9C9S0	Q9C9S0 arabidopsis
71	7	3.4	553	16	Q9RSN4	Q9RSN4 deinococcus
72	7	3.4	566	17	O97W79	O97W79 sulfolobus
73	7	3.4	589	5	O9B1M0	O9B1M0 leishmania
74	7	3.4	597	17	Q97UK4	Q97UK4 sulfolobus
75	7	3.4	617	16	O8Y1T3	O8Y1T3 raietonia s
76	7	3.4	658	16	O98IX1	O98IX1 rhizobium 1
77	7	3.4	722	4	O9Y223	O9Y223 homo sapien
78	7	3.4	722	11	O920P6	O920P6 mus musculu
79	7	3.4	722	11	O35826	O35826 rattus norv
80	7	3.4	722	11	O91W68	O91W68 mus musculu
81	7	3.4	743	16	O8XDV5	O8XDV5 escherichia
82	7	3.4	743	10	Q9ZVX5	Q9ZVX5 arabidopsis
83	7	3.4	788	16	Q9CJT9	Q9CJT9 pasteurella
84	7	3.4	791	10	O91G07	O91G07 oryza sativ
85	7	3.4	820	16	O8YKQ9	O8YKQ9 anabaena sp
86	7	3.4	840	5	O8SS52	O8SS52 dictyosteli
87	7	3.4	868	16	O8YKX9	O8YKX9 anabaena sp
88	7	3.4	873	10	O9S0J3	O9S0J3 arabidopsis
89	7	3.4	875	5	O8TVV7	O8TVV7 trypanosoma

90	3.4	885	16	Q8YV11	Q8YV11 anabaena sp	163	2.9	110	12	Q9WB24	Q9wb24 human calic
91	3.4	917	5	Q8TV6	Q8TV6 crypanosoma	164	2.9	110	12	Q9IPN8	Q9ipn8 normal-k-lik
92	3.4	1017	10	Q9M3H2	Q9m3h2 arabidopsis	165	2.9	110	12	Q9IPN2	Q9ipn2 normal-k-lik
93	3.4	1066	10	Q9S6ZC7	Q9s6zc7 arabidopsis	166	2.9	110	12	Q9WAL9	Q9wal9 human calic
94	3.4	1076	17	Q28994	Q28994 archaeoglob	167	2.9	111	2	Q47886	Q47886 frankia aln
95	3.4	1118	16	P73197	P73197 synechocyst	168	2.9	111	2	Q9KWM8	Q9kwm8 treponema m
96	3.4	1165	10	Q9CWM2	Q9cwm2 arabidopsis	169	2.9	111	12	Q88289	Q88289 small round
97	3.4	1239	16	Q9FBZ4	Q9fbz4 streptomyce	170	2.9	113	12	Q9QWMS	Q9qwms normal-k-lik
98	3.4	1633	16	Q8U634	Q8u634 agrobacteri	171	2.9	114	16	Q9CUI5	Q9cui5 lactococcus
99	3.4	2212	5	Q94657	Q94657 plasmodium	172	2.9	115	5	Q9V8F0	Q9v8f0 dirosophila
100	3.4	2364	2	Q46342	Q46342 clostridium	173	2.9	115	8	Q9TEG4	Q9teg4 cavia porce
101	3.4	2380	16	Q8XYE8	Q8xye8 ralistaonia s	174	2.9	117	2	Q9X954	Q9x954 streptomyce
102	3.4	2448	5	Q8SSS1	Q8sse1 dicystostell	175	2.9	118	4	Q9S217	Q9s217 homo sapien
103	3.4	5953	16	Q8XS39	Q8xs39 ralistaonia s	176	2.9	119	4	Q75484	Q75484 homo sapien
104	2.9	24	6	Q28780	Q28780 papio hamad	177	2.9	120	12	Q918X2	Q918x2
105	2.9	31	16	Q97NK1	Q97nk1 streptococc	178	2.9	120	12	Q918X1	Q918x1 human calic
106	2.9	38	2	Q47883	Q47883 frankia aln	179	2.9	121	11	Q55213	Q55213 rattus norv
107	2.9	43	6	Q9TSD0	Q9tsd0 macropus gl	180	2.9	121	11	Q68382	Q68382 human calic
108	2.9	56	9	Q37902	Q37902 xanthomonas	181	2.9	121	16	Q68383	Q68383 human calic
109	2.9	56	15	Q56181	Q56181 human immun	182	2.9	121	16	Q9K2P3	Q9k2p3 human calic
110	2.9	56	15	Q56182	Q56182 human immun	183	2.9	121	17	Q9HNM9	Q9hnm9 halobacteri
111	2.9	56	15	Q56183	Q56183 human immun	184	2.9	122	10	Q9SBV5	Q9sbv5
112	2.9	56	15	Q56184	Q56184 human immun	185	2.9	123	5	Q18565	Q18565 dirosophila
113	2.9	56	15	Q56185	Q56185 human immun	186	2.9	123	5	Q18566	Q18566 dirosophila
114	2.9	56	15	Q56187	Q56187 human immun	187	2.9	123	5	Q18567	Q18567 dirosophila
115	2.9	59	9	Q38657	Q38657 bacteriophage	188	2.9	123	5	Q18568	Q18568 dirosophila
116	2.9	61	10	Q8RV85	Q8rv85 pinus sylve	189	2.9	123	5	Q94847	Q94847 dirosophila
117	2.9	62	16	Q8YSN1	Q8ysn1 anabaena sp	190	2.9	123	5	Q94872	Q94872 dirosophila
118	2.9	66	12	Q9WAX6	Q9wax6 human calic	191	2.9	123	5	Q76346	Q76346 dirosophila
119	2.9	68	16	Q98HY9	Q98hy9 rhizobium l	192	2.9	123	5	Q76347	Q76347 dirosophila
120	2.9	72	12	Q8VAJ8	Q8vaj8 white spot	193	2.9	123	5	Q76348	Q76348 dirosophila
121	2.9	74	4	Q8TB04	Q8tb04 homo sapien	194	2.9	123	5	Q76349	Q76349 dirosophila
122	2.9	74	5	Q17236	Q17236 bombyx mori	195	2.9	124	12	Q9QW63	Q9qw63 normal-k-lik
123	2.9	78	5	Q95XU9	Q95xu9 caenothabdi	196	2.9	124	12	Q9QW69	Q9qw69 normal-k-lik
124	2.9	78	12	Q99B09	Q99b09 human calic	197	2.9	124	16	Q9KS62	Q9ks62
125	2.9	78	17	Q8ZVR3	Q8zvr3 pyrobaculum	198	2.9	125</			

236	6	2.9	147	2	Q8VT24	Q8Vt24 enterococu	309	6	2.9	194	16	Q8Z9C5	Q8Z9C5 salmonella
237	6	2.9	147	10	Q245S1	Q245S1 vitis vinif	310	6	2.9	195	16	Q8VK58	Q8VK58 mycobacteri
238	6	2.9	148	2	Q9X4E4	Q9X4E4 rhodobacter	311	6	2.9	197	16	Q98MP7	Q98MP7 rhizobium l
239	6	2.9	149	16	Q8VHR4	Q8Vhr4 bruceella me	312	6	2.9	198	16	Q8XR75	Q8XR75 ralestonia s
240	6	2.9	149	17	Q97JZ8	Q97Jz8 sulfolobus	313	6	2.9	200	2	Q9JW68	Q9JW68 anaplasmia m
241	6	2.9	151	16	Q9JWQ2	Q9JWq2 bacillus ha	314	6	2.9	200	5	Q9N6W9	Q9N6W9 dirosophila
242	6	2.9	151	16	Q574Z9	Q574Z9 pseudomonas	315	6	2.9	200	5	Q77Z59	Q77Z59 oryza sativ
243	6	2.9	151	16	Q9ZSH2	Q9Zsh2 rhizobium m	316	6	2.9	200	10	Q9LW65	Q9LW65 simian cyto
244	6	2.9	152	10	Q9LWZ5	Q9Lwz5 oryza sativ	317	6	2.9	201	12	Q986M5	Q986M5
245	6	2.9	153	4	Q9V5J4	Q9V5J4 homo sapien	318	6	2.9	203	10	Q9W2M1	Q9W2M1 dirosophila
246	6	2.9	154	5	Q8SRH9	Q8srh9 encephalito	319	6	2.9	203	10	Q6S0J8	Q6S0J8 oryza sativ
247	6	2.9	154	11	Q91YQ3	Q91Yq3 mus musculu	320	6	2.9	203	16	Q8XM45	Q8XM45 clostridium
248	6	2.9	154	16	Q9J1I8	Q9J1i8 staphylococ	321	6	2.9	204	16	Q9PLX2	Q9PLX2 campylobact
249	6	2.9	155	10	Q9LYE5	Q9LYe5 arabidopsis	322	6	2.9	204	16	Q9HUJ2	Q9HUJ2 pseudomonas
250	6	2.9	155	16	Q8YR71	Q8Yr71 anabaena sp	323	6	2.9	204	16	Q8U9Z3	Q8U9Z3 agrobacteri
251	6	2.9	155	16	Q9ZCJ8	Q9Zcj8 listeria mo	324	6	2.9	206	16	Q9Z4T0	Q9Z4T0 salmonella
252	6	2.9	156	10	Q245S2	Q245S2 vitis vinif	325	6	2.9	206	16	Q8RTV5	Q8RTV5 thermotaner
253	6	2.9	156	16	Q9L8K2	Q9Lbk2 pseudomonas	326	6	2.9	208	4	Q96KA8	Q96KA8 homo sapien
254	6	2.9	158	16	Q8VVB0	Q8Vvb0 anabaena sp	327	6	2.9	208	4	Q96KA8	Q96KA8 homo sapien
255	6	2.9	160	2	Q9L7V1	Q9L7v1 haemophilus	328	6	2.9	208	17	Q9HIG7	Q9HIG7 thermoplasm
256	6	2.9	162	5	Q8SRP4	Q8srp4 encephalito	329	6	2.9	210	10	Q9SSC7	Q9SSC7 arabidopsis
257	6	2.9	162	16	Q9ZVK7	Q9Zvk7 rhizobium m	330	6	2.9	210	12	Q9Q6P5	Q9Q6P5
258	6	2.9	165	2	Q4S403	Q4S403 bacillus su	331	6	2.9	210	17	Q8TW10	Q8TW10 methanopyru
259	6	2.9	166	5	Q9BHK2	Q9Bhk2 sabelia spa	332	6	2.9	211	10	Q9SMA1	Q9Sma1 oryza sativ
260	6	2.9	167	1	Q526O3	Q526O3 sulfolobus	333	6	2.9	212	2	Q9RQ12	Q9RQ12 bacteroides
261	6	2.9	167	1	Q8VLN8	Q8Vln8 pseudomonas	334	6	2.9	212	2	Q9EYB2	Q9EYB2 escherichia
262	6	2.9	167	4	Q9POE0	Q9Poe0 homo sapien	335	6	2.9	213	16	Q06552	Q06552 mycobacteri
263	6	2.9	167	10	Q9SR76	Q9sr76 arabidopsis	336	6	2.9	213	16	Q99ZD5	Q99ZD5 streptococc
264	6	2.9	167	13	Q919D0	Q919d0 oreochromis	337	6	2.9	213	16	Q97FUI	Q97FUI streptococc
265	6	2.9	167	17	Q8TXE9	Q8Txe9 methanopyru	338	6	2.9	213	17	Q58697	Q58697 pyrococcus
266	6	2.9	168	2	Q93GH2	Q93gh2 bacillus su	339	6	2.9	215	16	Q9A690	Q9A690 caulobacter
267	6	2.9	168	13	Q90YH6	Q90Yh6 oreochromis	340	6	2.9	215	16	Q97F16	Q97F16 clostridium
268	6	2.9	168	17	Q26967	Q26967 methanobact	341	6	2.9	215	16	Q8XSR1	Q8XSR1 escherichia
269	6	2.9	170	10	Q8RUX8	Q8Rux8 arabidopsis	342	6	2.9	216	2	Q9ZJZ2	Q9ZJZ2 burkholderi
270	6	2.9	170	16	Q9KTR0	Q9Ktr0 vibrio chol	343	6	2.9	216	2	Q9RKM7	Q9Rkm7 klebsiella
271	6	2.9	173	2	Q9FCP9	Q9Fcp9 phylllobacte	344	6	2.9	216	16	Q9CEB9	Q9CEB9 lactococcus
272	6	2.9	173	8	Q94YD3	Q94Yd3 harpaddon mi	345	6	2.9	216	16	Q98FP4	Q98FP4 rhizobium l
273	6	2.9	175	6	Q94TC1	Q94tc1 saurida und	346	6	2.9	216	16	Q8Y8V1	Q8Y8V1 listeria mo
274	6	2.9	175	6	Q9GLC1	Q9Glc1 sus scrofa	347	6	2.9	216	16	Q8Y1J3	Q8Y1J3 ralestonia s
275	6	2.9	177	2	Q9FCQ3	Q9Fcq3 phylllobacte	348	6	2.9	218	2	P94955	P94955 micrococcus
276	6	2.9	177	2	Q9FCU4	Q9Fcu4 mesorhizobi	349	6	2.9	218	5	Q8SV70	Q8SV70 encephalito
277	6	2.9	177	2	Q9FC55	Q9Fcs5 mesorhizobi	350	6	2.9	218	13	Q9YHA9	Q9Yha9 myxine glut
278	6	2.9	177	2	Q9FCT6	Q9Fct6 mesorhizobi	351	6	2.9	218	16	Q9ZDM9	Q9ZDM9 listeria in
279	6	2.9	177	2	Q9FCU0	Q9Fcu0 mesorhizobi	352	6	2.9	219	10	Q8RXM0	Q8RXM0 arabidopsis
280	6	2.9	177	2	Q9FCT8	Q9Fct8 rhizobium l	353	6	2.9	219	16	Q9ZMM3	Q9Zmm3 rhizobium m
281	6	2.9	177	16	Q9KFE6	Q9Kfe6 bacillus ha	354	6	2.9	220	16	Q9KXM2	Q9Kxm2 vibrio chol
282	6	2.9	178	2	Q938T4	Q938t4 uncultured	355	6	2.9	220	16	Q9A273	Q9A273 caulobacter
283	6	2.9	178	16	Q9FC23	Q9Fcz3 streptomyc	356	6	2.9	221	4	Q9HCF6	Q9Hcf6 homo sapien
284	6	2.9	179	5	Q9W0E9	Q9Woe9 dirosophila	357	6	2.9	221	16	Q9HYF3	Q9HYF3 pseudomonas
285	6	2.9	179	16	Q26O45	Q26O45 helicobacte	358	6	2.9	221	17	Q9Y9A8	Q9Y9A8 aeropyrum p
286	6	2.9	181	16	Q8UHB8	Q8Uhb8 agrobacteri	359	6	2.9	222	16	Q92YV7	Q92Yv7 rhizobium m
287	6	2.9	182	2	Q9ZNA9	Q9Zna9 streptomyc	360	6	2.9	223	2	Q9S6L2	Q9S6L2 corynebacte
288	6	2.9	184	10	Q94EC1	Q94ec1 oryza sativ	361	6	2.9	223	3	Q96UK2	Q96UK2 zygosacchar
289	6	2.9	184	16	Q9RDM6	Q9Rdm6 streptomyc	362	6	2.9	224	3	Q03J22	Q03J22 saccharomyc
290	6	2.9	186	5	Q19714	Q19714 caenorhabdi	363	6	2.9	224	10	Q9SKX3	Q9SKX3 arabidopsis
291	6	2.9	186	16	Q9KQ38	Q9Kq38 vibrio chol	364	6	2.9	225	2	Q05758	Q05758 mycobacteri
292	6	2.9	186	17	Q26678	Q26678 methanobact	365	6	2.9	225	3	Q9UYE1	Q9UYe1 zygosacchar
293	6	2.9	187	4	Q96DT9	Q96dt9 homo sapien	366	6	2.9	226	4	Q9H0E0	Q9H0E0 homo sapien
294	6	2.9	187	11	Q9WUZ5	Q9Wuz5 mus musculu	367	6	2.9	226	16	Q9UX96	Q9UX96 naiseeria m
295	6	2.9	187	16	Q8ZSJ6	Q8Zsj6 anabaena sp	368	6	2.9	227	11	Q63430	Q63430 rattus norv
296	6	2.9	188	5	Q9VBE9	Q9Vbe9 dirosophila	369	6	2.9	227	11	Q91XO6	Q91Xg6 mus musculu
297	6	2.9	188	13	Q90J30	Q90J30 coturnix co	370	6	2.9	228	16	Q8YDD5	Q8Ydd5 bruceella me
298	6	2.9	189	2	Q935X5	Q935X5 gynecococc	371	6	2.9	229	17	Q8ZUD0	Q8Zud0 pyrobaculum
299	6	2.9	189	17	Q8TWM5	Q8Twm5 mechanopyru	372	6	2.9	230	16	Q9A0S4	Q9A0S4 streptococc
300	6	2.9	190	2	Q938T5	Q938t5 uncultured	373	6	2.9	231	5	Q9VN35	Q9VN35 dirosophila
301	6	2.9	190	9	Q9AKG7	Q9akg7 bacterioph	374	6	2.9	232	5	Q69245	Q69245 bradyrhizob
302	6	2.9	190	16	Q9CEA6	Q9cea6 lactococcus	375	6	2.9	232	5	Q76350	Q76350 dirosophila
303	6	2.9	190	16	Q97DM5	Q97dm5 clostridium	376	6	2.9	232	16	Q8X8Z9	Q8X8Z9 escherichia
304	6	2.9	191	5	Q9V8E8	Q9v8e8 dirosophila	377	6	2.9	232	16	Q8Y1Y7	Q8Y1Y7 ralestonia s
305	6	2.9	191	16	Q9A6D4	Q9a6d4 caulobacter	378	6	2.9	233	13	Q9PSG2	Q9PSg2 gallus gall
306	6	2.9	191	17	Q8TPE5	Q8Tpe5 mechanosarc	379	6	2.9	234	10	Q8SSS9	Q8SSS9 arabidopsis
307	6	2.9	192	10	Q947U7	Q947u7 oryza sativ	380	6	2.9	234	16	Q8YS22	Q8Ys22 anabaena sp
308	6	2.9	192	16	Q9HTU3	Q9htu3 pseudomonas	381	6	2.9	235	2	Q9RLC4	Q9RLC4 pseudomonas

382	6	2.9	235	16	09FBR1	09FBI1 streptomyc	455	6	2.9	256	16	053757	053757 mycobacteri
383	6	2.9	236	4	09HUV2	09HUV2 homo sapien	456	6	2.9	257	10	09SEK8	09SEK8 glycine max
384	6	2.9	236	4	09HUV7	09HUV7 homo sapien	457	6	2.9	257	16	09ZAT5	09ZAT5 listeria in
385	6	2.9	236	5	0811Y7	0811Y7 dictyosteli	458	6	2.9	257	16	08Y6G4	08Y6G4 listeria mo
386	6	2.9	236	16	0914W0	0914W0 pseudomonas	459	6	2.9	257	16	P71568	P71568 mycobacteri
387	6	2.9	237	5	09UBT1	09UBT1 drosophila	460	6	2.9	257	16	09ZC12	09ZC12 streptomyc
388	6	2.9	237	5	09UBT4	09UBT4 drosophila	461	6	2.9	257	16	09ZC12	09ZC12 streptomyc
389	6	2.9	237	5	09UBT3	09UBT3 drosophila	462	6	2.9	257	16	09ZC12	09ZC12 streptomyc
390	6	2.9	237	5	09UBT2	09UBT2 drosophila	463	6	2.9	257	16	09ZC12	09ZC12 streptomyc
391	6	2.9	237	5	09UBS9	09UBS9 drosophila	464	6	2.9	258	2	093UM1	093UM1 synchococ
392	6	2.9	237	5	09UBS5	09UBS5 drosophila	465	6	2.9	258	12	002483	002483 bovine hep
393	6	2.9	237	5	09UBS5	09UBS5 drosophila	466	6	2.9	258	16	09Z7N5	09Z7N5 caulobacter
394	6	2.9	237	5	09UBS4	09UBS4 drosophila	467	6	2.9	258	16	09Z7N5	09Z7N5 caulobacter
395	6	2.9	237	5	09UBS3	09UBS3 drosophila	468	6	2.9	259	12	0987G2	0987G2 western equ
396	6	2.9	237	5	09UBR7	09UBR7 lordiphosa	469	6	2.9	259	13	0987G4	0987G4 anguilla ja
397	6	2.9	237	16	025636	025636 helicobacte	470	6	2.9	259	13	0987G4	0987G4 anguilla ja
398	6	2.9	237	16	08YSG3	08YSG3 anabaena sp	471	6	2.9	259	13	0987G4	0987G4 anguilla ja
399	6	2.9	238	9	09X9R4	09X9R4 streptomyc	472	6	2.9	259	13	0987G2	0987G2 anguilla ja
400	6	2.9	239	9	091UX3	091UX3 bacterioph	473	6	2.9	260	16	0987G1	0987G1 anguilla ja
401	6	2.9	239	16	09PPA7	09PPA7 campylobact	474	6	2.9	260	16	09ZNL4	09ZNL4 anabaena sp
402	6	2.9	239	16	09A9K5	09A9K5 caulobacter	475	6	2.9	260	16	09ZNL4	09ZNL4 anabaena sp
403	6	2.9	239	16	098803	098803 rhizobium 1	476	6	2.9	261	2	09RNM3	09RNM3 aeropyrum p
404	6	2.9	240	5	09N4B3	09N4B3 caenorhabdi	477	6	2.9	261	2	09RNM3	09RNM3 aeropyrum p
405	6	2.9	240	10	0948U1	0948U1 magnolia pr	478	6	2.9	261	2	09RNM3	09RNM3 aeropyrum p
406	6	2.9	240	11	09DBL5	09DBL5 mus musculu	479	6	2.9	261	2	09RNM3	09RNM3 aeropyrum p
407	6	2.9	240	11	09R3N8	09R3N8 mus musculu	480	6	2.9	261	2	09RNM3	09RNM3 aeropyrum p
408	6	2.9	240	11	09R3N8	09R3N8 mus musculu	481	6	2.9	261	2	09RNM3	09RNM3 aeropyrum p
409	6	2.9	240	16	0930R6	0930R6 rhizobium m	482	6	2.9	262	10	09ATL7	09ATL7 zeax mays (m
410	6	2.9	242	16	09K8R9	09K8R9 bacillus ha	483	6	2.9	264	2	053023	053023 zeax mays (m
411	6	2.9	243	16	097G19	097G19 clostridium	484	6	2.9	264	10	09FWV6	09FWV6 zeax mays (m
412	6	2.9	243	16	08U8Q5	08U8Q5 agrobacteri	485	6	2.9	265	16	08ZWN2	08ZWN2 zeax mays (m
413	6	2.9	244	16	066269	066269 erythromicr	486	6	2.9	265	16	09P1Z3	09P1Z3 zeax mays (m
414	6	2.9	244	2	09R7K1	09R7K1 erythrobact	487	6	2.9	266	2	095387	095387 zeax mays (m
415	6	2.9	244	2	0875R4	0875R4 staphylococ	488	6	2.9	266	2	095387	095387 zeax mays (m
416	6	2.9	244	2	051345	051345 pseudomonas	489	6	2.9	266	10	09ATL7	09ATL7 zeax mays (m
417	6	2.9	245	2	066272	066272 erythrobact	490	6	2.9	267	10	004589	004589 arabisidopsi
418	6	2.9	245	2	066278	066278 erythrobact	491	6	2.9	267	12	097D48	097D48 semiki for
419	6	2.9	245	2	082987	082987 erythrobact	492	6	2.9	267	12	097D48	097D48 semiki for
420	6	2.9	245	2	09ZNM7	09ZNM7 erythrobact	493	6	2.9	267	12	097D48	097D48 semiki for
421	6	2.9	245	2	09ZNM7	09ZNM7 erythrobact	494	6	2.9	267	12	097D48	097D48 semiki for
422	6	2.9	245	2	09ZNM7	09ZNM7 erythrobact	495	6	2.9	267	12	097D48	097D48 semiki for
423	6	2.9	245	2	09ZNM7	09ZNM7 erythrobact	496	6	2.9	267	12	097D48	097D48 semiki for
424	6	2.9	245	2	09ZNM7	09ZNM7 erythrobact	497	6	2.9	267	12	097D48	097D48 semiki for
425	6	2.9	246	2	066276	066276 blaetomonas	498	6	2.9	269	2	09LH43	09LH43 mycobacteri
426	6	2.9	247	16	P71851	P71851 mycobacteri	499	6	2.9	269	2	09LH43	09LH43 mycobacteri
427	6	2.9	248	5	09XW20	09XW20 caenorhabdi	500	6	2.9	269	2	09LH43	09LH43 mycobacteri
428	6	2.9	248	16	09HJ33	09HJ33 pseudomonas	501	6	2.9	269	2	09LH43	09LH43 mycobacteri
429	6	2.9	249	2	082989	082989 erythrobact	502	6	2.9	269	2	09LH43	09LH43 mycobacteri
430	6	2.9	249	2	09ZEX6	09ZEX6 buchiera ap	503	6	2.9	269	10	09ZV57	09ZV57 streptococ
431	6	2.9	249	17	08UOB5	08UOB5 pyrococcus	504	6	2.9	269	16	P96378	P96378 mycobacteri
432	6	2.9	250	10	08R187	08R187 arabisidopsi	505	6	2.9	269	16	P96378	P96378 mycobacteri
433	6	2.9	250	16	09A7R8	09A7R8 caulobacter	506	6	2.9	270	2	047757	047757 enterococ
434	6	2.9	251	13	09W621	09W621 cymops pyr	507	6	2.9	270	2	047757	047757 enterococ
435	6	2.9	251	16	025588	025588 helicobacte	508	6	2.9	270	2	047757	047757 enterococ
436	6	2.9	252	10	09FWL2	09FWL2 oryza sativ	509	6	2.9	270	11	08QZY6	08QZY6 mus musculu
437	6	2.9	253	4	09HJ34	09HJ34 homo sapien	510	6	2.9	270	16	08XV14	08XV14 anabaena ap
438	6	2.9	253	5	09XU41	09XU41 caenorhabdi	511	6	2.9	270	16	08XV14	08XV14 anabaena ap
439	6	2.9	253	5	016093	016093 drosophila	512	6	2.9	271	2	093HC5	093HC5 ralestoma s
440	6	2.9	253	10	094118	094118 oryza sativ	513	6	2.9	271	17	08TNI3	08TNI3 streptomyc
441	6	2.9	253	16	08ZG56	08ZG56 versinia pe	514	6	2.9	272	10	098501	098501 guillardia
442	6	2.9	254	5	022883	022883 caenorhabdi	515	6	2.9	273	4	09NX07	09NX07 homo sapien
443	6	2.9	254	5	027595	027595 drosophila	516	6	2.9	273	10	08S1G4	08S1G4 oryza sativ
444	6	2.9	254	5	096925	096925 drosophila	517	6	2.9	274	16	08ZRL8	08ZRL8 salmonella
445	6	2.9	254	5	096925	096925 drosophila	518	6	2.9	274	16	08ZRL8	08ZRL8 salmonella
446	6	2.9	254	5	096925	096925 drosophila	519	6	2.9	275	3	093852	093852 mycobacteri
447	6	2.9	254	5	096925	096925 drosophila	520	6	2.9	275	3	093852	093852 mycobacteri
448	6	2.9	254	5	096925	096925 drosophila	521	6	2.9	275	3	093852	093852 mycobacteri
449	6	2.9	254	5	096925	096925 drosophila	522	6	2.9	275	3	093852	093852 mycobacteri
450	6	2.9	254	5	096925	096925 drosophila	523	6	2.9	275	3	093852	093852 mycobacteri
451	6	2.9	254	5	096925	096925 drosophila	524	6	2.9	275	3	093852	093852 mycobacteri
452	6	2.9	254	5	096925	096925 drosophila	525	6	2.9	275	3	093852	093852 mycobacteri
453	6	2.9	254	5	096925	096925 drosophila	526	6	2.9	275	3	093852	093852 mycobacteri
454	6	2.9	254	5	096925	096925 drosophila	527	6	2.9	275	3	093852	093852 mycobacteri

528	6	2.9	279	16	Q9A1H4	Q9A1H4 streptococc	601	6	2.9	302	12	Q8V310	Q8V310 swinepox vi
529	6	2.9	279	16	Q984S7	Q984S7 rhizobium 1	602	6	2.9	302	16	Q8XBL6	Q8XBL6 escherichia
530	6	2.9	279	16	Q8Z3X1	Q8Z3X1 salmonella	603	6	2.9	302	16	Q915G0	Q915G0 salmonella
531	6	2.9	280	2	Q9RDX8	Q9RDX8 legionella	604	6	2.9	303	2	Q91731	Q91731 streptomyces
532	6	2.9	280	2	Q32396	Q32396 erythrobact	605	6	2.9	304	11	Q9R165	Q9R165 mus musculus
533	6	2.9	280	10	Q91LZK9	Q91LZK9 arabidopsis	606	6	2.9	304	16	Q92AH5	Q92AH5 listeria in
534	6	2.9	280	16	Q8YCH3	Q8YCH3 bruceella me	607	6	2.9	304	16	Q8Y667	Q8Y667 listeria mo
535	6	2.9	281	2	Q8VN96	Q8VN96 helicobacte	608	6	2.9	305	2	Q8KWC1	Q8KWC1 agrobacteri
536	6	2.9	281	2	Q8VN94	Q8VN94 helicobacte	609	6	2.9	305	16	Q8A936	Q8A936 caulobacter
537	6	2.9	281	2	Q60250	Q60250 staphylococ	610	6	2.9	305	16	Q8ZDP3	Q8ZDP3 yersinia pe
538	6	2.9	281	2	Q51699	Q51699 staphylococ	611	6	2.9	308	11	Q8D9N7	Q8D9N7 mus musculu
539	6	2.9	281	16	Q9AC80	Q9AC80 staphylococ	612	6	2.9	308	16	Q9CJY9	Q9CJY9 pasteurella
540	6	2.9	281	16	Q98FS2	Q98FS2 rhizobium 1	613	6	2.9	308	16	Q8BNW9	Q8BNW9 rhizobium 1
541	6	2.9	283	4	Q92V10	Q92V10 rhizobium m	614	6	2.9	308	16	Q8Z8G8	Q8Z8G8 salmonella
542	6	2.9	283	4	Q9NY06	Q9NY06 homo sapien	615	6	2.9	309	2	Q51272	Q51272 neisseria m
543	6	2.9	284	2	Q9Z3H4	Q9Z3H4 symbiococc	616	6	2.9	309	16	Q99XT9	Q99XT9 streptococ
544	6	2.9	284	10	Q9W563	Q9W563 vitis vinif	617	6	2.9	309	16	Q8UDR0	Q8UDR0 agrobacteri
545	6	2.9	284	10	Q9S2Q7	Q9S2Q7 arabidopsis	618	6	2.9	310	2	Q9EXF1	Q9EXF1 listeria mo
546	6	2.9	284	16	Q9RFP4	Q9RFP4 vibrio chol	619	6	2.9	310	10	Q8VZ31	Q8VZ31 arabidopsis
547	6	2.9	285	5	Q9VFP6	Q9VFP6 drosoephila	620	6	2.9	310	17	Q8E545	Q8E545 archaeoglob
548	6	2.9	285	16	Q8XHG8	Q8XHG8 clostridium	621	6	2.9	311	16	Q9HTC9	Q9HTC9 pseudomonas
549	6	2.9	286	2	Q30544	Q30544 agrobacteri	622	6	2.9	311	16	Q98LFP1	Q98LFP1 rhizobium 1
550	6	2.9	286	2	Q9RCD6	Q9RCD6 agrobacteri	623	6	2.9	311	16	Q98FL8	Q98FL8 rhizobium 1
551	6	2.9	286	2	Q49628	Q49628 mycobacteri	624	6	2.9	312	5	Q8WPH1	Q8WPH1 bomblyx mori
552	6	2.9	286	5	Q21152	Q21152 caenorhabdi	625	6	2.9	312	16	Q9KS22	Q9KS22 vibrio chol
553	6	2.9	286	10	Q81023	Q81023 arabidopsis	626	6	2.9	312	16	Q97R65	Q97R65 streptococ
554	6	2.9	286	16	Q9KDL5	Q9KDL5 bacillus ha	627	6	2.9	313	2	Q9S4W0	Q9S4W0 salmonella
555	6	2.9	286	16	Q8U604	Q8U604 agrobacteri	628	6	2.9	313	2	Q9RNU4	Q9RNU4 zymomonas m
556	6	2.9	287	2	Q9KX46	Q9KX46 clostridium	629	6	2.9	313	15	Q80754	Q80754 human t-lym
557	6	2.9	287	16	Q912V3	Q912V3 pseudomonas	630	6	2.9	313	15	Q80771	Q80771 human t-lym
558	6	2.9	287	16	Q8RSP3	Q8RSP3 thermoanaer	631	6	2.9	313	16	Q8UB67	Q8UB67 agrobacteri
559	6	2.9	287	17	Q8TL85	Q8TL85 methanosaer	632	6	2.9	314	4	Q9Y3Y6	Q9Y3Y6 homo sapien
560	6	2.9	288	5	Q9NE08	Q9NE08 caenorhabdi	633	6	2.9	314	10	Q40489	Q40489 nicotiana t
561	6	2.9	288	16	Q9BNW2	Q9BNW2 rhizobium 1	634	6	2.9	314	11	Q8VFX1	Q8VFX1 mus musculu
562	6	2.9	288	16	Q97QX4	Q97QX4 streptococ	635	6	2.9	314	16	Q8Z9H8	Q8Z9H8 salmonella
563	6	2.9	289	10	Q9M1X0	Q9M1X0 arabidopsis	636	6	2.9	314	16	Q8UI03	Q8UI03 agrobacteri
564	6	2.9	289	12	Q83879	Q83879 norwayk vir	637	6	2.9	315	11	Q8VEW4	Q8VEW4 mus musculu
565	6	2.9	289	12	Q66295	Q66295 calicivir	638	6	2.9	316	2	Q9RH80	Q9RH80 neisseria m
566	6	2.9	289	16	Q9RXY6	Q9RXY6 deinococcus	639	6	2.9	316	2	Q9RH79	Q9RH79 neisseria m
567	6	2.9	290	2	Q06453	Q06453 streptococ	640	6	2.9	316	2	Q66068	Q66068 neisseria m
568	6	2.9	290	2	Q53114	Q53114 rhodobacter	641	6	2.9	316	5	Q9V8P0	Q9V8P0 drosoephila
569	6	2.9	290	17	Q9YDT4	Q9YDT4 aeropyrum p	642	6	2.9	316	10	Q9FUC6	Q9FUC6 hevea bras
570	6	2.9	291	5	Q17843	Q17843 caenorhabdi	643	6	2.9	316	16	Q9KMU0	Q9KMU0 vibrio chol
571	6	2.9	291	5	Q8SWA1	Q8SWA1 encephaliti	644	6	2.9	316	16	Q9HWT8	Q9HWT8 pseudomonas
572	6	2.9	291	13	Q9DPF41	Q9DPF41 brachydantio	645	6	2.9	316	16	Q8XXK4	Q8XXK4 ralsionia s
573	6	2.9	291	17	Q8ZVH8	Q8ZVH8 pyrobaculum	646	6	2.9	317	2	Q9Z3B6	Q9Z3B6 neisseria m
574	6	2.9	292	10	Q9LH15	Q9LH15 arabidopsis	647	6	2.9	317	17	Q9H8B8	Q9H8B8 halobacteri
575	6	2.9	292	10	Q8W5C4	Q8W5C4 oryza sativ	648	6	2.9	318	2	Q87449	Q87449 neisseria m
576	6	2.9	292	13	Q9PSG1	Q9PSG1 gallus gall	649	6	2.9	320	2	Q9ACD8	Q9ACD8 vibrio salm
577	6	2.9	292	16	Q9XAH8	Q9XAH8 streptomyc	650	6	2.9	320	17	Q8U1V1	Q8U1V1
578	6	2.9	292	16	Q8XO81	Q8XO81 ralsionia s	651	6	2.9	321	16	Q9ZTH1	Q9ZTH1 rhizobium m
579	6	2.9	293	5	Q9Y1Z3	Q9Y1Z3 hydra magni	652	6	2.9	321	16	Q9Z5Y0	Q9Z5Y0 rhizobium m
580	6	2.9	293	5	Q16423	Q16423 caenorhabdi	653	6	2.9	322	2	Q9Z1Z5	Q9Z1Z5 neisseria m
581	6	2.9	294	2	Q9XCL4	Q9XCL4 bartonella	654	6	2.9	322	2	Q87447	Q87447 neisseria m
582	6	2.9	294	2	Q9FP926	Q9FP926 pasteurella	655	6	2.9	322	2	Q87448	Q87448 neisseria m
583	6	2.9	294	5	Q8SVQ1	Q8SVQ1 encephaliti	656	6	2.9	322	2	Q66061	Q66061 neisseria m
584	6	2.9	294	12	Q68290	Q68290 human calic	657	6	2.9	322	2	Q66062	Q66062 neisseria m
585	6	2.9	294	16	Q9CK73	Q9CK73 pasteurella	658	6	2.9	322	2	Q66058	Q66058 neisseria m
586	6	2.9	296	16	Q97KZ8	Q97KZ8 clostridium	659	6	2.9	322	2	Q66058	Q66058 neisseria m
587	6	2.9	296	16	Q8XUM0	Q8XUM0 ralsionia s	660	6	2.9	322	2	Q51273	Q51273 neisseria m
588	6	2.9	297	5	Q9N4B4	Q9N4B4 caenorhabdi	661	6	2.9	322	5	Q17846	Q17846 caenorhabdi
589	6	2.9	297	16	Q8ZM77	Q8ZM77 salmonella	662	6	2.9	322	10	Q9SHN5	Q9SHN5 arabidopsis
590	6	2.9	297	16	Q8UAC7	Q8UAC7 agrobacteri	663	6	2.9	323	2	Q990Z5	Q990Z5 enterobacte
591	6	2.9	298	16	Q8XO15	Q8XO15 ralsionia s	664	6	2.9	323	2	Q53900	Q53900 streptococ
592	6	2.9	298	16	Q9A1D5	Q9A1D5 streptococ	665	6	2.9	323	2	Q66057	Q66057 neisseria m
593	6	2.9	299	12	Q91959	Q91959 psammecchinu	666	6	2.9	323	2	Q66060	Q66060 neisseria m
594	6	2.9	299	12	Q99CZ1	Q99CZ1 bovine herp	667	6	2.9	323	2	Q66063	Q66063 neisseria m
595	6	2.9	301	16	Q9KPI8	Q9KPI8 vibrio chol	668	6	2.9	323	2	Q66067	Q66067 neisseria m
596	6	2.9	301	16	P71064	P71064 bacillus su	669	6	2.9	323	2	Q51276	Q51276 neisseria m
597	6	2.9	301	16	Q97FK3	Q97FK3 clostridium	670	6	2.9	323	2	Q70078	Q70078 neisseria m
598	6	2.9	302	2	Q24625	Q24625 bacillus eh	671	6	2.9	323	2	Q51274	Q51274 neisseria m
599	6	2.9	302	12	Q9YWP6	Q9YWP6 colitivirus	672	6	2.9	323	2	Q51275	Q51275 neisseria m
600	6	2.9	302	12	Q9YWP6	Q9YWP6 colitivirus	673	6	2.9	323	5	Q9N365	Q9N365 caenorhabdi

674	6	2.9	323	16	Q98J14	Q98J14 rhizobium 1	747	6	2.9	355	16	Q8R831	Q8R831 thermoaer
675	6	2.9	324	9	Q8SCK1	Q8SCK1 bacterioph	748	6	2.9	356	2	Q9ZHX7	Q9ZHX7 pasteurella
676	6	2.9	324	12	Q9D1C1	Q9D1C1 pan troglod	749	6	2.9	356	2	Q8V869	Q8V869 borrelia he
677	6	2.9	324	16	Q9AKA0	Q9AKA0 streptomyc	750	6	2.9	356	17	Q9HS17	Q9HS17 halobacteri
678	6	2.9	325	2	Q9Z5X8	Q9Z5X8 fireaureia s	751	6	2.9	357	16	Q98FQ1	Q98FQ1 rhizobium 1
679	6	2.9	325	17	Q8UJ12	Q8UJ12 pyrococcus	752	6	2.9	357	16	Q8XC40	Q8XC40 escherichia
680	6	2.9	326	11	Q91XD8	Q91XD8 mus musculu	753	6	2.9	358	4	Q96HF8	Q96HF8 homo sapien
681	6	2.9	326	16	Q8YUW9	Q8YUW9 anabaena sp	754	6	2.9	358	5	Q9GTH5	Q9GTH5 caenorhabdi
682	6	2.9	327	2	Q9ZERT	Q9ZERT xanthobacte	755	6	2.9	360	2	Q56047	Q56047 streptococc
683	6	2.9	327	16	Q986B1	Q986B1 rhizobium 1	756	6	2.9	361	5	Q97220	Q97220 leishmania
684	6	2.9	328	2	Q9X5G8	Q9X5G8 streptomyc	757	6	2.9	361	16	Q8ZCW2	Q8ZCW2 yersinia pe
685	6	2.9	328	4	Q9NHA0	Q9NHA0 streptomyc	758	6	2.9	363	4	Q96F69	Q96F69 homo sapien
686	6	2.9	328	6	Q9N085	Q9N085 macaca fasc	759	6	2.9	364	10	Q23334	Q23334 arabisdopsis
687	6	2.9	329	16	Q9CJY9	Q9CJY9 pasteurella	760	6	2.9	364	12	Q92543	Q92543 hepatitis c
688	6	2.9	330	16	Q9GLX0	Q9GLX0 bos taurus	761	6	2.9	365	10	Q91B45	Q91B45 pan rhadno
689	6	2.9	330	16	Q9KXU4	Q9KXU4 vibrio chol	762	6	2.9	365	5	Q20846	Q20846 caenorhabdi
690	6	2.9	331	10	Q81475	Q81475 arabisdopsis	763	6	2.9	365	10	Q40103	Q40103 kalanchoe b
691	6	2.9	331	17	Q97UH3	Q97UH3 sulfolobus	764	6	2.9	365	10	Q40104	Q40104 kalanchoe b
692	6	2.9	331	17	Q8Z294	Q8Z294 pyrobaculum	765	6	2.9	365	16	Q9X0A4	Q9X0A4 thermotoga
693	6	2.9	333	4	Q96CH1	Q96CH1 homo sapien	766	6	2.9	367	2	Q9WM77	Q9WM77 rhizobium 1
694	6	2.9	333	12	Q66232	Q66232 calicivirus	767	6	2.9	367	12	Q72468	Q72468 nelson bay
695	6	2.9	333	16	Q910X1	Q910X1 pseudomonas	768	6	2.9	367	2	Q9WM77	Q9WM77 rhizobium 1
696	6	2.9	334	4	Q8WUW6	Q8WUW6 homo sapien	769	6	2.9	368	5	Q9VDY6	Q9VDY6 rhizobium 1
697	6	2.9	334	16	Q8YTF7	Q8YTF7 anabaena sp	770	6	2.9	368	10	Q9MT75	Q9MT75 homo sapien
698	6	2.9	335	4	Q9WME3	Q9WME3 agrobacteri	771	6	2.9	369	17	Q9YF57	Q9YF57 drosophila
699	6	2.9	335	4	Q9H2F9	Q9H2F9 homo sapien	772	6	2.9	368	4	Q8WZ58	Q8WZ58 drosophila
700	6	2.9	336	3	Q42703	Q42703 candida par	773	6	2.9	369	10	Q9MT75	Q9MT75 malomonas
701	6	2.9	336	16	Q986C5	Q986C5 rhizobium 1	774	6	2.9	371	2	Q9AL6	Q9AL6 aeropyrum p
702	6	2.9	336	16	Q92975	Q92975 archaeoglob	775	6	2.9	372	11	Q9CRC4	Q9CRC4 mus musculu
703	6	2.9	337	13	Q90Y27	Q90Y27 sphyaena 1	776	6	2.9	372	16	Q8ZPR0	Q8ZPR0 salmonella
704	6	2.9	337	16	Q92VQ2	Q92VQ2 rhizobium m	777	6	2.9	372	2	Q9AL6	Q9AL6 salmonella
705	6	2.9	337	16	Q92VM2	Q92VM2 rhizobium m	778	6	2.9	373	10	Q9SSH0	Q9SSH0 arabisdopsis
706	6	2.9	338	17	Q8TU08	Q8TU08 methanosarc	779	6	2.9	373	10	Q94GVO	Q94GVO oryza sativ
707	6	2.9	339	5	P91321	P91321 caenorhabdi	780	6	2.9	374	16	Q8UAM6	Q8UAM6 agrobacteri
708	6	2.9	339	10	Q04907	Q04907 arabisdopsis	781	6	2.9	374	16	Q8XJ11	Q8XJ11 caulobacter
709	6	2.9	339	16	Q9K048	Q9K048 vibrio chol	782	6	2.9	375	16	Q915B3	Q915B3 escherichia
710	6	2.9	340	2	Q44324	Q44324 agrobacteri	783	6	2.9	376	4	Q9BR60	Q9BR60 homo sapien
711	6	2.9	340	5	Q26744	Q26744 trypanosoma	784	6	2.9	376	11	Q91Z25	Q91Z25 mus musculu
712	6	2.9	340	5	Q62367	Q62367 caenorhabdi	785	6	2.9	376	16	Q983B2	Q983B2 rhizobium 1
713	6	2.9	341	16	Q981I5	Q981I5 rhizobium 1	786	6	2.9	377	16	Q8UAM6	Q8UAM6 mycobacteri
714	6	2.9	341	16	Q9X4P8	Q9X4P8 salmonella 1	787	6	2.9	377	16	Q8UAM6	Q8UAM6 tydha pomon
715	6	2.9	341	16	Q983B6	Q983B6 rhizobium 1	788	6	2.9	378	12	Q91E09	Q91E09 trachemys s
716	6	2.9	343	5	Q9W3B2	Q9W3B2 drosophila	789	6	2.9	378	13	Q91E09	Q91E09 escherichia
717	6	2.9	343	16	Q9KXJ4	Q9KXJ4 vibrio chol	790	6	2.9	378	16	Q69724	Q69724 caenorhabdi
718	6	2.9	344	2	P91928	P91928 uncultured	791	6	2.9	379	4	Q9NV52	Q9NV52 yersinia pe
719	6	2.9	345	2	Q8RTT9	Q8RTT9 drosophila	792	6	2.9	379	16	Q9K5X8	Q9K5X8 notropis at
720	6	2.9	346	5	Q8T0F8	Q8T0F8 arabisdopsis	793	6	2.9	380	16	Q9KT24	Q9KT24 mus musculu
721	6	2.9	346	10	Q944K7	Q944K7 neisseria m	794	6	2.9	381	2	Q9EVA2	Q9EVA2 acinetobact
722	6	2.9	346	16	Q9JVR8	Q9JVR8 yersinia pe	795	6	2.9	381	16	Q914S2	Q914S2 pseudomonas
723	6	2.9	346	16	Q8ZJF2	Q8ZJF2 yersinia pe	796	6	2.9	382	2	P94791	P94791 flavobacter
724	6	2.9	346	16	Q8ZDS8	Q8ZDS8 yersinia pe	797	6	2.9	382	10	Q9STN6	Q9STN6 caenorhabdi
725	6	2.9	347	10	Q9ZU77	Q9ZU77 arabisdopsis	798	6	2.9	382	16	Q8YPL6	Q8YPL6 clostridium
726	6	2.9	348	8	Q99948	Q99948 notropis at	799	6	2.9	383	16	Q8XN89	Q8XN89 ralistonia s
727	6	2.9	348	10	Q8SK40	Q8SK40 ctenophorus	800	6	2.9	383	16	Q9GDM2	Q9GDM2 caenorhabdi
728	6	2.9	348	12	Q93877	Q93877 glycine max	801	6	2.9	384	16	Q8XG65	Q8XG65 ralistonia s
729	6	2.9	348	12	Q9J5U2	Q9J5U2 o hyong-nyo	802	6	2.9	384	10	Q40492	Q40492 nicotiana t
730	6	2.9	348	16	Q9KOP0	Q9KOP0 neisseria m	803	6	2.9	385	5	Q96830	Q96830 rattus sp.
731	6	2.9	349	1	Q49598	Q49598 methanopyru	804	6	2.9	386	16	Q9A2P4	Q9A2P4 caulobacter
732	6	2.9	349	16	Q9RW33	Q9RW33 deinococcus	805	6	2.9	388	16	Q9N308	Q9N308 caenorhabdi
733	6	2.9	350	4	Q96M49	Q96M49 homo sapien	806	6	2.9	389	10	Q93AK7	Q93AK7 aphanirocc
734	6	2.9	350	16	Q96JL6	Q96JL6 rhizobium 1	807	6	2.9	389	12	Q91B44	Q91B44 haematococ
735	6	2.9	351	5	Q9GN06	Q9GN06 babesia big	808	6	2.9	389	10	Q9AUV0	Q9AUV0 pan rhadno
736	6	2.9	351	16	Q8XJV8	Q8XJV8 escherichia	809	6	2.9	389	12	Q91B43	Q91B43 pan rhadno
737	6	2.9	351	17	Q8ZTY7	Q8ZTY7 pyrobaculum	810	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
738	6	2.9	352	10	Q9X082	Q9X082 neurospora	811	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
739	6	2.9	352	10	Q9ZP99	Q9ZP99 hevea bras	812	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
740	6	2.9	352	12	Q93280	Q93280 equine herp	813	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
741	6	2.9	353	13	Q8OF85	Q8OF85 xenopus lae	814	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
742	6	2.9	353	16	Q99UE7	Q99UE7 scaphylococ	815	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
743	6	2.9	353	16	Q99UC8	Q99UC8 streptomyc	816	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
744	6	2.9	353	16	Q99UC8	Q99UC8 streptomyc	817	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
745	6	2.9	355	8	Q21316	Q21316 zea mays (m	818	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1
746	6	2.9	355	10	Q9FMH5	Q9FMH5 arabisdopsis	819	6	2.9	389	16	Q961X0	Q961X0 rhizobium 1

820	6	2.9	389	16	Q92810	Q92810 listeria in	893	6	2.9	418	16	Q98LY1	Q98LY1 rhizobium 1
821	6	2.9	389	16	Q9Y4K9	Q9Y4K9 listeria mo	894	6	2.9	419	10	Q9SGT3	Q9SGT3 arabisidopsis
822	6	2.9	390	5	Q962K0	Q962K0 plasmidum	895	6	2.9	419	16	Q9K1K5	Q9K1K5 natsieria m
823	6	2.9	390	16	Q9L189	Q9L189 streptomyc	896	6	2.9	419	16	Q9JWZ0	Q9JWZ0 natsieria m
824	6	2.9	391	11	Q9DABA	Q9DABA mus musculu	897	6	2.9	420	16	Q9AD12	Q9AD12 streptomyc
825	6	2.9	393	10	Q9C9P8	Q9C9P8 arabisidopsis	898	6	2.9	421	10	Q9FX57	Q9FX57 arabisidopsis
826	6	2.9	394	2	Q9R8E5	Q9R8E5 burkholderi	899	6	2.9	421	16	Q34528	Q34528 bacillus su
827	6	2.9	395	17	Q29908	Q29908 archaeoglob	900	6	2.9	421	16	Q9S1N5	Q9S1N5 streptomyc
828	6	2.9	398	2	Q36689	Q36689 rhizobium 1	901	6	2.9	421	17	Q8TJ12	Q8TJ12 methanarc
829	6	2.9	398	4	Q9Y406	Q9Y406 homo sapien	902	6	2.9	422	9	Q9XJ12	Q9XJ12 bacterioph
830	6	2.9	399	2	Q69789	Q69789 rhodococcus	903	6	2.9	422	10	Q9FV44	Q9FV44 tegetes ere
831	6	2.9	399	16	Q91458	Q91458 pseudomonas	904	6	2.9	422	16	Q9CLD8	Q9CLD8 pasteurella
832	6	2.9	399	16	Q86832	Q86832 streptomyc	905	6	2.9	422	16	Q8Y266	Q8Y266 ralsionia s
833	6	2.9	400	4	Q9N196	Q9N196 homo sapien	906	6	2.9	422	16	Q9KXB5	Q9KXB5 escherichia
834	6	2.9	400	16	Q9R8S2	Q9R8S2 deinococcus	907	6	2.9	423	16	Q927H4	Q927H4 listeria in
835	6	2.9	401	2	Q935X2	Q935X2 acinetobact	908	6	2.9	423	16	Q8Y4J2	Q8Y4J2 listeria mo
836	6	2.9	401	4	Q9BU00	Q9BU00 homo sapien	909	6	2.9	425	16	Q92GA7	Q92GA7 rickettsia
837	6	2.9	401	16	Q97TL4	Q97TL4 clostridium	910	6	2.9	425	16	Q8RDJ5	Q8RDJ5 thermoaer
838	6	2.9	402	2	Q8RR21	Q8RR21 streptomyc	911	6	2.9	426	2	Q9RLJ0	Q9RLJ0 pseudomonas
839	6	2.9	402	3	Q94388	Q94388 schizosacch	912	6	2.9	426	11	Q55082	Q55082 mus musculu
840	6	2.9	402	5	Q62319	Q62319 caenorhabdi	913	6	2.9	426	16	Q9XBR2	Q9XBR2 bacillus ha
841	6	2.9	402	12	Q9DVZ7	Q9DVZ7 plutella xy	914	6	2.9	427	3	Q93983	Q93983 hansenia a
842	6	2.9	402	16	Q8Y2V3	Q8Y2V3 ralsionia s	915	6	2.9	427	16	Q92629	Q92629 listeria in
843	6	2.9	402	16	Q8RFT3	Q8RFT3 fubobacteri	916	6	2.9	427	16	Q8Y3T4	Q8Y3T4 listeria mo
844	6	2.9	403	3	Q13633	Q13633 schizosacch	917	6	2.9	428	16	Q910W0	Q910W0 pseudomonas
845	6	2.9	403	10	Q9FGK6	Q9FGK6 arabisidopsis	918	6	2.9	429	11	Q8R1S6	Q8R1S6 mus musculu
846	6	2.9	403	16	Q8YQU4	Q8YQU4 arabidops	919	6	2.9	430	16	Q9A5U3	Q9A5U3 caulobacter
847	6	2.9	404	5	Q45394	Q45394 caenorhabdi	920	6	2.9	430	16	Q9A267	Q9A267 caulobacter
848	6	2.9	404	10	Q22791	Q22791 arabisidopsis	921	6	2.9	431	2	Q9XBR9	Q9XBR9 zymomonas m
849	6	2.9	404	16	Q92V81	Q92V81 rhizobium m	922	6	2.9	431	10	Q80902	Q80902 arabisidopsis
850	6	2.9	404	16	P94988	P94988 mycobacteri	923	6	2.9	431	16	Q9Y660	Q9Y660 streptococ
851	6	2.9	404	16	Q9RXZ9	Q9RXZ9 deinococcus	924	6	2.9	433	5	Q19701	Q19701 caenorhabdi
852	6	2.9	404	17	Q9YFV9	Q9YFV9 aeropyrum p	925	6	2.9	433	5	Q9Y961	Q9Y961 drosophila
853	6	2.9	404	17	Q979C7	Q979C7 thermoplasma	926	6	2.9	434	16	Q9KXK7	Q9KXK7 vibrio chol
854	6	2.9	405	10	Q65263	Q65263 arabisidopsis	927	6	2.9	434	16	Q8ZIC8	Q8ZIC8 yersinia pe
855	6	2.9	406	2	Q9K2H2	Q9K2H2 mycobacteri	928	6	2.9	435	11	Q99KX6	Q99KX6 mus musculu
856	6	2.9	406	10	Q92WJ6	Q92WJ6 zea mays (m	929	6	2.9	436	2	Q9FA01	Q9FA01 athrobacte
857	6	2.9	406	16	Q99TP6	Q99TP6 staphylococ	930	6	2.9	436	10	Q9XGM7	Q9XGM7 citrus para
858	6	2.9	407	11	Q8ROA7	Q8ROA7 mus musculu	931	6	2.9	436	10	Q9M608	Q9M608 citrus unsh
859	6	2.9	407	16	Q9HX99	Q9HX99 pseudomonas	932	6	2.9	436	12	Q9DY07	Q9DY07 grapevine f
860	6	2.9	407	16	Q9FXW3	Q9FXW3 streptomyc	933	6	2.9	436	12	Q9DY06	Q9DY06 grapevine f
861	6	2.9	410	2	Q9ANW7	Q9ANW7 streptococ	934	6	2.9	436	12	Q9DY05	Q9DY05 grapevine f
862	6	2.9	410	2	Q9L445	Q9L445 streptococ	935	6	2.9	436	16	Q9KLO8	Q9KLO8 vibrio chol
863	6	2.9	410	2	Q9L445	Q9L445 streptococ	936	6	2.9	437	3	Q12007	Q12007 saccharomyc
864	6	2.9	410	2	Q9L446	Q9L446 streptococ	937	6	2.9	437	10	Q9SPN3	Q9SPN3 arabisidopsis
865	6	2.9	410	4	Q8WZ89	Q8WZ89 homo sapien	938	6	2.9	437	13	Q42316	Q42316 cyprinus ca
866	6	2.9	410	4	Q8WZ88	Q8WZ88 homo sapien	939	6	2.9	439	2	Q66167	Q66167 agrobacteri
867	6	2.9	410	6	Q9GL51	Q9GL51 sus scrofa	940	6	2.9	439	4	Q9UD03	Q9UD03 homo sapien
868	6	2.9	410	10	Q94126	Q94126 oryza sativ	941	6	2.9	439	13	Q98TS5	Q98TS5 xenopus lae
869	6	2.9	410	13	Q9PTR5	Q9PTR5 gallus gall	942	6	2.9	439	16	Q9R706	Q9R706 agrobacteri
870	6	2.9	410	13	Q90ZL4	Q90ZL4 xenopus lae	943	6	2.9	439	16	Q9RSF6	Q9RSF6 deinococcus
871	6	2.9	410	16	Q9L388	Q9L388 streptococ	944	6	2.9	440	10	Q9M6E2	Q9M6E2 taxus cuspi
872	6	2.9	411	5	Q96698	Q96698 drosophila	945	6	2.9	440	10	Q8W169	Q8W169 taxus bacca
873	6	2.9	411	10	Q944G3	Q944G3 hevea brasl	946	6	2.9	440	11	Q922R8	Q922R8 mus musculu
874	6	2.9	411	16	Q92JH8	Q92JH8 rickettsia	947	6	2.9	441	16	Q970X1	Q970X1 streptococ
875	6	2.9	412	3	Q00569	Q00569 hansenia a	948	6	2.9	442	2	Q9F7P1	Q9F7P1 uncultured
876	6	2.9	413	3	Q94657	Q94657 schizosacch	949	6	2.9	442	16	Q8XW88	Q8XW88 ralsionia s
877	6	2.9	413	16	Q9PF76	Q9PF76 xyliella feb	950	6	2.9	442	17	Q8TZJ0	Q8TZJ0 pyrococcus
878	6	2.9	414	2	Q9AMW9	Q9AMW9 bradyrhizob	951	6	2.9	443	10	Q9FJN0	Q9FJN0 arabisidopsis
879	6	2.9	414	5	P95421	P95421 pseudomonas	952	6	2.9	443	16	Q9PN09	Q9PN09 campylobact
880	6	2.9	414	5	Q25308	Q25308 leishmania	953	6	2.9	443	16	Q9WYL1	Q9WYL1 thermotoga
881	6	2.9	414	10	Q9SJM4	Q9SJM4 arabisidopsis	954	6	2.9	444	16	Q8XGV4	Q8XGV4 ralsionia s
882	6	2.9	414	12	Q65303	Q65303 turkey rhin	955	6	2.9	444	17	Q57949	Q57949 pyrococcus
883	6	2.9	414	12	Q65305	Q65305 turkey rhin	956	6	2.9	444	17	Q8U4G2	Q8U4G2 pyrococcus
884	6	2.9	414	16	Q910Y9	Q910Y9 pseudomonas	957	6	2.9	445	13	Q98841	Q98841 anguilla an
885	6	2.9	414	16	Q95EW2	Q95EW2 streptomyc	958	6	2.9	445	13	Q98842	Q98842 anguilla an
886	6	2.9	416	4	Q8TDT5	Q8TDT5 homo sapien	959	6	2.9	446	8	Q8WND7	Q8WND7 bos taurus
887	6	2.9	416	16	Q92UP4	Q92UP4 rhizobium m	960	6	2.9	446	8	Q9MNX3	Q9MNX3 paxurus lon
888	6	2.9	416	16	Q8Y135	Q8Y135 ralsionia s	961	6	2.9	446	8	Q9WGD2	Q9WGD2 penaeus mon
889	6	2.9	417	2	Q45037	Q45037 borrelia bu	962	6	2.9	446	8	Q8WA95	Q8WA95 narceus ann
890	6	2.9	417	16	Q9KSF5	Q9KSF5 vibrio chol	963	6	2.9	446	13	Q42315	Q42315 cyprinus ca
891	6	2.9	417	16	Q9KP78	Q9KP78 vibrio chol	964	6	2.9	446	17	Q9V278	Q9V278 pyrococcus
892	6	2.9	417	16	Q50899	Q50899 borrelia bu	965	6	2.9	447	3	Q42889	Q42889 schizosacch

```

966 6 2.9 447 10 Q9FFQ0 Q9ffq0 arabidopsis
967 6 2.9 448 12 Q9V6J3 Q9v6j3 californi
968 6 2.9 448 16 Q9JRB0 Q9jrb0 neisseria m
969 6 2.9 450 2 Q528G9 Q528g9 rnzodum 1
970 6 2.9 450 16 Q9HT50 Q9ht50 pseudomonas
971 6 2.9 450 16 Q9ZHC5 Q9zhc5 yersinia pe
972 6 2.9 450 16 Q8ZG48 Q8zgj8 yersinia pe
973 6 2.9 453 16 Q8ZK26 Q8znk6 salmonella
974 6 2.9 453 16 Q8ZAL3 Q8zal3 salmonella
975 6 2.9 454 16 Q9ZGL6 Q9zgl6 listeria in
976 6 2.9 455 5 Q9GQW9 Q9gqn9 schistosoma
977 6 2.9 455 16 Q9F2W4 Q9f2w4 streptomyce
978 6 2.9 457 5 Q8T4G9 Q8t4g9 drosophila
979 6 2.9 457 11 Q9JUN5 Q9jnj5 mus muscula
980 6 2.9 457 11 Q9EOV8 Q9eov8 ratu8 norv
981 6 2.9 457 12 Q8S046 Q8s046 pseudorabie
982 6 2.9 457 17 Q8ZU73 Q8zu73 pyrobaculum
983 6 2.9 458 10 Q9LJ78 Q9lj78 arabidopsis
984 6 2.9 458 11 Q9LWM9 Q9lwm9 mus musculu
985 6 2.9 459 4 Q7S108 Q7s108 homo sapien
986 6 2.9 459 16 Q5S594 Q5s594 synchocyst
987 6 2.9 460 8 Q9ASH2 Q9ash2 gasterocyst
988 6 2.9 460 16 Q916R7 Q916r7 pseudomonas
989 6 2.9 461 2 Q8VN93 Q8vn93 helicobacte
990 6 2.9 461 2 Q8VN92 Q8vn92 helicobacte
991 6 2.9 462 10 Q9T018 Q9t018 arabidopsis
992 6 2.9 462 16 Q9ZCU8 Q9zcu8 listeria in
993 6 2.9 462 16 Q8YBK3 Q8ybk3 brucella me
994 6 2.9 463 2 Q9EY60 Q9ey60 amycolatops
995 6 2.9 463 5 Q16751 Q16751 caenorhabdi
996 6 2.9 463 12 Q8S045 Q8s045 pseudorabie
997 6 2.9 464 10 Q9SE01 Q9se01 glycine max
998 6 2.9 464 10 Q94EA9 Q94ea9 cryza sativ
999 6 2.9 465 16 Q8U7K8 Q8u7k8 agrobacteri
1000 6 2.9 465 17 Q8TS27 Q8ts27 metanosarc

```

ALIGNMENTS

```

RESULT 1
ID Q9DCV4 PRELIMINARY; PRT; 305 AA.
AC Q9DCV4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 2410005016Rik protein.
GN 2410005016Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085560; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa T.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,

```

```

RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002442; BAB22104.1; -.
DR MGD; MGI:1913552; 2410005016Rik.
DR InterPro; IPR001440; TPR.
SQ SEQUENCE 305 AA; 34998 MW; 98A7855743A61C7A CRC64;

```

```

Query Match 5.4%; Score 11; DB 11; Length 305;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 94 FEKATLNPXD 104
Db 189 FEKATLNPXD 199

```

```

RESULT 2
ID Q9CWS6 PRELIMINARY; PRT; 305 AA.
AC Q9CWS6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 2410005016Rik protein.
GN 2410005016Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMERYONIC STEM CELLS;
RX MEDLINE=21085560; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa T.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010421; BAB26924.1; -.
DR MGD; MGI:1913552; 2410005016Rik.
DR InterPro; IPR001440; TPR.
SQ SEQUENCE 305 AA; 35000 MW; 692F34973A5FA8D CRC64;

```

```

Query Match 5.4%; Score 11; DB 11; Length 305;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 94 FEKATLNPXD 104
Db 189 FEKATLNPXD 199

```

```

RESULT 3
ID Q8R166 PRELIMINARY; PRT; 305 AA.
AC Q8R166;

```

DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE RIKEN_CDNA_241005016 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025158; AA025158.1; -
SQ SEQUENCE 305 AA; 35030 MW; DD2A248637ACA618 CRC64;

Query Match
Best Local Similarity 100.0%; Score 11; DB 11; Length 305;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 FEKAIEINPKD 104
DB 189 FEKAIEINPKD 199

RESULT 4
ID Q9Y398 PRELIMINARY; PRT; 314 AA.
AC Q9Y398;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DE CGI-90 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of Novel Human Genes Evolutionarily Conserved in
Caenorhabditis elegans by Comparative Proteomics.";
RL Genome Res. 10:703-713 (2000).
DR EMBL; AF151848; AAD34085.1; -
SQ SEQUENCE 314 AA; 35716 MW; 230A93514AC2BAF1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 11; DB 4; Length 314;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 FEKAIEINPKD 104
DB 194 FEKAIEINPKD 204

RESULT 5
ID Q96DB5 PRELIMINARY; PRT; 314 AA.
AC Q96DB5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Similar to CGI-90 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009671; AA009671.1; -

SQ SEQUENCE 314 AA; 35808 MW; 4976011FDF3C389 CRC64;
Query Match
Best Local Similarity 100.0%; Score 11; DB 4; Length 314;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 FEKAIEINPKD 104
DB 194 FEKAIEINPKD 204

RESULT 6
ID O26186 PRELIMINARY; PRT; 379 AA.
AC O26186;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE O-linked GlcNAc transferase.
GN MTH83.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxId=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lunn W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Saefer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155 (1997).
DR EMBL; AE000799; AAB84589.1; -
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 10.
DR SMART; SM00028; TPR; 10.
DR Transferase; Complete proteome.
SQ SEQUENCE 379 AA; 43975 MW; 476138ABE9EB83E7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 10; DB 17; Length 379;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 FEKAIEINPK 103
DB 175 FEKAIEINPK 184

RESULT 7
ID O04646 PRELIMINARY; PRT; 212 AA.
AC O04646;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE A_TMO21B04.14 protein (Hypohectical 23.3 kDa protein).
GN A_TMO21B04.14 OR A75G27280.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Dante M., Wamsley P., Gibson A.;
RT "The sequence of A. thaliana TMO21B04.";

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Wash-U;
 RT "The A. thaliana Genome Sequencing Project."
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Waterston R.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF007271; AAB61076.1; -;
 DR EMBL; AY081569; AAM10131.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 212 AA; 23318 MW; DAA63E482F35B89D CRC64;
 Query Match 3.9%; Score 8; DB 10; Length 212;
 Best Local Similarity 100.0%; Pred. No. 5.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 12 AMGLVLSA 19
 Db 72 AMGLVLSA 79
 RESULT 8
 Q9DDR1 PRELIMINARY; PRT; 397 AA.
 AC Q9DDR1;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-JUN-2001 (TrEMBLrel. 16, Last sequence update)
 DE Thymotrophin-releasing hormone receptor 1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenoipodidae; Xenopus.
 NCBI_TaxID=8355;
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Dammann O., Harder S., Buck F., Richter D., Bruhn T.O.;
 RT "Differential Expression of Two Distinct Thymotrophin Releasing-Hormone
 Receptors Cloned from Xenopus laevis Brain."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF305620; AAC40849.1; -;
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PRINTS; PR00237; GPCRHHODPSN.
 DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; UNKNOWN_1.
 DR PROSITE; PS0262; G_PROTEIN_RECP_F1_2; 1.
 KW Receptor.
 SQ SEQUENCE 397 AA; 44970 MW; 919436AFAD7AD7D CRC64;

Query Match 3.9%; Score 8; DB 13; Length 397;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 83 DATHQTKA 90
 Db 237 DATHQTKA 244
 RESULT 9
 Q9MOM6 PRELIMINARY; PRT; 515 AA.
 AC Q9MOM6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE Hypothetical 58.6 kDa protein.
 GN A74G17220.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL161546; CAB78725.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 515 AA; 58640 MW; 764E57652801F326 CRC64;
 Query Match 3.9%; Score 8; DB 10; Length 515;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 171 SALERALI 178
 Db 159 SALERALI 166
 RESULT 10
 O23565 PRELIMINARY; PRT; 550 AA.
 AC O23565;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Myosin II heavy chain like protein.
 GN D14645C.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA Bevan M., Strikema W., Murphy G., Wambutt R., Pohl T., Terry N.,
 RA Kreis M., Kavanagh T., Ehtian K.D., Rieger M., James R.,
 RA Pridmorenech P., Hatzopoulos P., Oberwaler B., Duesterhoft A.,
 RA Jones J., Palme K., Ansgore W., Delseny M., Bancroft I., Mewes H.W.,
 RA Schueller C., Chaiwatiz N.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z97343; CAB10503.1; -;
 SQ SEQUENCE 550 AA; 62424 MW; 0D89E75C02C0A0C2 CRC64;
 Query Match 3.9%; Score 8; DB 10; Length 550;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 SALERALI 178
 |||||
 Db 159 SALERALI 166

RESULT 11

O9Y7T4 PRELIMINARY; PRT; 830 AA.

AC O9Y7T4; PRELIMINARY; PRT; 830 AA.

DT 01-NOV-1999 (TEMBLrel. 12, Created)

DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)

DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)

DE Putative serine/threonine-protein kinase C63.08C (EC 2.7.1.-).

GN SPC63.08C.

OS Schizosaccharomyces pombe (Pission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC AFG/UNC-51/URK1 SUBFAMILY.

EMBL AL049522.CAB40012.1; -

DR InterPro: IPR000719; Euk_Pkinase.

DR InterPro: IPR002290; Ser_thr_kinase.

DR InterPro: IPR001245; Tyr_kinase.

DR Pfam: PF00069; pkinase; 1.

DR PRINTS: PR00109; TYRKINASE.

DR ProDom: PD000001; Euk_pkinase; 1.

DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.

KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;

KW ATP-binding.

FT DOMAIN 14 307 PROTEIN KINASE.

FT NP_BIND 20 28 ATP (BY SIMILARITY).

FT BINDING 43 43 ATP (BY SIMILARITY).

FT ACT_SITE 157 157 BY SIMILARITY.

FT ACT_SITE 157 157 BY SIMILARITY.

SO SEQUENCE 830 AA; 91584 MW; 98B76C502BD83425 CRC64;

Query Match 3.9%; Score 8; DB 3; Length 830;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 KSALERALI 177

Db 541 KSALERALI 548

RESULT 12

O9QMW7 PRELIMINARY; PRT; 2154 AA.

AC O9QMW7; PRELIMINARY; PRT; 2154 AA.

DT 01-MAY-2000 (TEMBLrel. 13, Created)

DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)

DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)

DE Non-erythrocyte beta spectrin.

GN SPN82.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN=C57 BLACK/6;

RA Cat T., Yu P., Mishra B., Monga P.S.P., Mishra L.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF017112; AAD01616.1; -

DR HSP; 001082; 1BKR.

DR MGD; MGI:98388; Spn2.

DR InterPro: IPR002106; AACRNA_ligaseII.

DR InterPro: IPR001589; Actbind actin.

DR InterPro: IPR001715; Calponin-like.

DR InterPro: IPR002017; Spectrin.

DR Pfam: PF00307; CH; 2.

DR Pfam: PF00435; spectrin; 17.

DR SMART: SM00033; CH; 2.

DR SMART: SM00150; SPEC; 16.

DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.

DR PROSITE: PS00019; ACTININ_1.

DR PROSITE: PS00020; ACTININ_2; 1.

DR PROSITE: PS50021; CH; 2.

SO SEQUENCE 2154 AA; 250928 MW; 872CPEC9C332152C CRC64;

Query Match 3.9%; Score 8; DB 11; Length 2154;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 QTKAORLF 94

Db 1357 QTKAORLF 1364

RESULT 13

O30275 PRELIMINARY; PRT; 78 AA.

AC O30275; PRELIMINARY; PRT; 78 AA.

DT 01-JAN-1998 (TEMBLrel. 05, Created)

DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)

DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)

DE Hypothetical protein AF2396.

GN AF2396.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

OC Archaeoglobaceae; Archaeoglobus.

OX NCBI_TaxID=2234;

RP [1]

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,

RA Richardson D.L., Kervatage A.R., Graham D.E., Kyriades N.C.,

RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,

RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,

RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,

RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,

RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Weese C.R.,

RA Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-

reducing archaeon Archaeoglobus fulgidus.";

RL Nature 390:364-370 (1997).

DR EMBL: AE001111; AAB91275.1; -

KW TIGR; AF2396;

SO SEQUENCE 78 AA; 7975 MW; ADBF974647878936 CRC64;

Query Match 3.4%; Score 7; DB 17; Length 78;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 138 EGRVVAI 144

Db 14 EGRVVAI 20

RESULT 14

O8UUK9 PRELIMINARY; PRT; 107 AA.

ID O8UUK9; PRELIMINARY; PRT; 107 AA.

AC O8UUK9; PRELIMINARY; PRT; 107 AA.

DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
 DE Hypothetical protein Atus467.
 GN Atus467 OR AGR_PAT 689.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Plasmid AT.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OC NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Moo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. St.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Kreppan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Neeser E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58.";
 RT Science 294:2317-2323 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Houmelo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmelo K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Garsen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58.";
 RT Science 294:2323-2328 (2001).
 DR EMBL; AE007916; AAL46154.1; ALT_INIT.
 DR EMBL; AE008968; AAL46154.1; ALT_INIT.
 DR EMBL; AE007916; AAK90842.1; -
 KW Hypothetical protein; Plasmid; Complete proteome.
 SQ SEQUENCE 107 AA; 11700 MW; 958B4B07F2D8544 CRC64;

Query Match
 Best Local Similarity 3.4%; Score 7; DB 16; Length 107;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 170 KSALEA 176
 Db 27 KSALEA 33

RESULT 15
 O8TYX6 PRELIMINARY; PRT; 120 AA.
 AC O8TYX6;
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
 DE Uncharacterized protein conserved in archaea.
 GN MK0165.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 OC Methanopyrus.
 OC NCBI_TaxID=3320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
 RA Stcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Natche D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Stetter K.O.,
 RA Malyn A.G., Koonin E.V., Kozaykin S.A.;

RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
 DR EMBL; AE010316; AAM01382.1; -
 KW Complete proteome.
 SQ SEQUENCE 120 AA; 13331 MW; EE4703EC5A1FF5B CRC64;

Query Match
 Best Local Similarity 3.4%; Score 7; DB 17; Length 120;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 AKSALER 175
 Db 12 AKSALER 18

RESULT 16
 ID O97610 PRELIMINARY; PRT; 121 AA.
 AC O97610;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DE Hypothetical protein ST0208.
 GN ST0208.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OC NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yangsi M., Nishimura M., Yamagishi A.,
 RA Oshida T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermophilic RT Crenarchaeon, Sulfolobus tokodaii strain7.";
 RT DNA Res. 8:123-140 (2001).
 DR EMBL; AP000981; BAB5167.1; -
 DR InterPro: IPR002833; UPR0059.
 DR Pfam: PF01981; UPR0059; 1.
 DR ProDom: PD010667; UPR0059; 1.
 DR TIGRPFAMs: TIGR00283; UPR0059; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 121 AA; 13264 MW; EBE370D6A036FCB CRC64;

Query Match
 Best Local Similarity 3.4%; Score 7; DB 17; Length 121;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 EKAIELN 101
 Db 73 EKAIELN 79

RESULT 17
 ID O67408 PRELIMINARY; PRT; 131 AA.
 AC O67408;
 DT 01-AUG-1998 (T-EMBLrel. 07, Created)
 DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
 DE Hypothetical protein A0_1409.
 GN A0_1409.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 OC NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=VF5;
RX MEDLINE=9819666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).
DR EMBL: AE000739; AAC07377.1; -
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 131 AA; 15774 MW; 49E428E0D8580709 CRC64;

Query Match 3.4%; Score 7; DB 16; Length 131;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 AYRTLAK 76
DB 79 AYRTLAK 85
|||||
- - - - -

RESULT 18
QYX322 PRELIMINARY; PRT; 149 AA.
ID QYX322;
AC QYX322;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE PX01-52.
OS Bacillus anthracis.
OC Plasmid virulence plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STERNE;
RX MEDLINE=99445483; PubMed=10515943;
RA Okinka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,
RA Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,
RA Martinez Y., Rieke D., Svensson R., Jackson P.J.;
RT "Sequence and organization of PX01, the large Bacillus anthracis
RT plasmid harboring the anthrax toxin genes";
RL J. Bacteriol. 181:6509-6515(1999).
DR EMBL: AF065404; AAD32356.1; -
KW Plasmid.
SQ SEQUENCE 149 AA; 16366 MW; D17DC9D8B1FB56F6 CRC64;

Query Match 3.4%; Score 7; DB 2; Length 149;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 KSALERA 176
DB 35 KSALERA 41
|||||
- - - - -

RESULT 19
QY4475 PRELIMINARY; PRT; 156 AA.
ID QY4475;
AC QY4475;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE MUC.
OS Azotobacter vinelandii.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Azotobacter.
OX NCBI_TaxID=354;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=UM 136;
RX MEDLINE=96178940; PubMed=8606151;
RA Martinez-Salazar J.M., Moreno S., Najera R., Boucher J.C., Espin G.,
RA Soberon-Chavez G., Detric V.;
RT "Characterization of the genes coding for the putative sigma factor
RT Algu and its regulators MucA, MucB, MucC, and MucD in Azotobacter
RT vinelandii and evaluation of their roles in alginate biosynthesis";
RL J. Bacteriol. 178:1800-1808(1996).
DR EMBL: U30799; AAB01512.1; -
SQ SEQUENCE 156 AA; 17295 MW; 0C7849E5322A1C52 CRC64;

Query Match 3.4%; Score 7; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 GRVVAIE 145
DB 6 GRVVAIE 12
|||||
- - - - -

RESULT 20
P70952 PRELIMINARY; PRT; 164 AA.
ID P70952;
AC P70952; O08145;
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Hypothetical protein yltz.
GN yltz.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M168 CA;
RX MEDLINE=97177785; PubMed=9025291;
RA Levine A., Vannier F., Roche B., Autret S., Mavel D., Seror S.J.;
RT "A 10.3 kbp segment from nprB to argU at the 102 degrees region of the
RT Bacillus subtilis chromosome";
RL Microbiology 143:175-177(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98015415; PubMed=9353931;
RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and cItG (289
RT degrees) in Bacillus subtilis";
RL Microbiology 143:3305-3308(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.U., Commerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine J.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Pario V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
 RA Totsu V., Uchiyama S., Vanderbol M., Vannier F., Vassartotti A.,
 RA Viart A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
 RT *subtilis*,"
 RL Nature 390:249-256(1997).
 DR EMBL; Z79580; CAB01841.1; -
 DR EMBL; Y09476; CAA70637.1; -
 DR EMBL; Z99109; CAB12959.1; -
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 164 AA; 18353 MW; 2AD4D6970FA74D27 CRC64;

Query Match 3.4%; Score 7; DB 16; Length 164;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 LSGALIV 127
 Db 35 LSGALIV 41

RESULT 21

Q97AS8 PRELIMINARY; PRT; 165 AA.
 AC Q97AS8;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DN Hypothetical protein T070731.
 GN T070731 OR T060737684.
 OS Thermoplasma volcanum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=50339;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GS1 / DSM 4299 / JCM 9571;
 RX MEDLINE=20570466; PubMed=11121031;
 RA Kawashima T., Anano N., Koike H., Makino S.-I., Higuchi S.,
 RA Kawashima T., Anano N., Koike H., Makino S.-I., Higuchi S.,
 RA Nunoishi T., Yamamoto Y., Aizawa K., Yamazaki M., Kanemori K., Kawamoto T.,
 RT "Archaeal adaptation to higher temperatures revealed by genomic
 RT sequence of *Thermoplasma volcanum*,"
 RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
 DR EMBL; AP000993; BAB59873.1; -
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 165 AA; 19437 MW; 330439F902C2B2D0 CRC64;

Query Match 3.4%; Score 7; DB 17; Length 165;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 191 DKLLSDY 197
 Db 11 DKLLSDY 17

RESULT 22

Q9M500 PRELIMINARY; PRT; 176 AA.
 AC Q9M500;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE NBS-LRR-like protein (Fragment).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Euphorbiaceae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. YUXI 2; TISSUE=LEAF;
 RA Yang Q.Z., Wang P.W., Wang Q., Li J.R., Yan B., Huang X.Q.,
 RT "Rice Resistance Gene Analog Sequence,"
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF224493; AAF35363.1; -
 DR InterPro; IPR00767; Disease_resist.
 DR InterPro; IPR002182; NB-ARC.
 DR Pfam; PF00931; NB-ARC; 1.
 DR PRINTS; PR00364; DISEASERSIST.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 176 AA; 20172 MW; 9BE87E0422EC2FC9 CRC64;

Query Match 3.4%; Score 7; DB 10; Length 176;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 HDEIKKS 189
 Db 52 HDEIKKS 58

RESULT 23

Q98B31 PRELIMINARY; PRT; 183 AA.
 AC Q98B31;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Outer membrane lipoprotein.
 GN MUR5751.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kameo T., Nakamura Y., Sato S., Asamiizu E., Kato T., Sasamoto S.,
 RA Kameo T., Nakamura Y., Sato S., Asamiizu E., Kato T., Sasamoto S.,
 RA Kishida Y., Kiyokawa C., Kohara W., Matsunoto M., Matsuno A.,
 RA Wochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takenuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT *Mesorhizobium loti*,"
 RL DNA Res. 7:331-336(2000).
 DR EMBL; AP003007; BAB52141.1; -
 DR InterPro; IPR000566; Lipocin_cyFABP.
 DR Pfam; PF00061; Lipocalin; 1.
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
 KM Lipoprotein; Complete proteome.
 SQ SEQUENCE 183 AA; 20989 MW; 84BCFA7540AEP9C CRC64;

Query Match 3.4%; Score 7; DB 16; Length 183;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 KPSRAIG 136
 Db 76 KPSRAIG 82

RESULT 24

Q9R484 PRELIMINARY; PRT; 190 AA.
 AC Q9R484;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE ylb.
 GN ylb.
 OS Agrobacterium tumefaciens.
 OC plasmid Ti.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 NCBI_TaxID=358;
 OX
 RN
 RP SEQUENCE FROM N.A.
 RA Winans S.C., Zhu J., Oger P.M., Schrammeyer B., Hooykaas P.J.,
 RA Farrand S.K.;
 RT "Octopine-type Ti plasmid sequence."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF242881; AAD50316.1; -
 DR HSSP; P03012; 28SL.
 DR InterPro; IPR001822; Recombinase.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00239; resolvase; 1.
 DR PROSITE; PS00398; RECOMBINASES_2; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
 DR plasmid.
 KW
 SQ SEQUENCE 190 AA; 20793 MW; 3AF10B7F7977F85 CRC64;

Query Match 3.4%; Score 7; DB 2; Length 190;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 173 LERALI 179
 DB 95 LERALI 101

RESULT 25
 O9A170 PRELIMINARY; PRT; 193 AA.
 AC O9A170;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE Hypothetical protein A11838.
 GN A11838.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 NCBI_TaxID=103690;
 OX
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=PHB11;
 RA Blanco-Rivero A. Jr., Leganes F. Sr., Fernandez-Pinas F. Sr.;
 RT "Analysis of a tns-generated mutant of the cyanobacterium Anabaena sp.
 RT PCC7120 denoted as strain PHB11 that is impaired in growth at high
 RT pH."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasaamoto S.,
 RA Watanabe A., Iritaguchi W., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120."
 RL DNA Ref. 8:205-213(2001).
 DR EMBL; AF239799; AAK15078.1; -
 DR EMBL; AF003587; BAB73537.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 193 AA; 21313 MW; 7A3DAB375AA9D31B CRC64;

Query Match 3.4%; Score 7; DB 16; Length 193;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 LERALI 178
 DB 128 LERALI 134

RESULT 26
 O64536 PRELIMINARY; PRT; 204 AA.
 AC O64536;
 DT 01-AUG-1998 (TReMBLrel. 07, Created)
 DT 01-AUG-1998 (TReMBLrel. 07, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE YUP8H12R.23 protein.
 GN YUP8H12R.23.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 OX
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A., Vysotskaia V.S., Osborne B.I., Schwartz J.R.,
 RA Federpsiel N.A., Kwan A., Toriumi M., Yu G., Oji, O., Araujo R.,
 RA Chung E., Dewar K., Dietrich F., Ecker J.R., Marzilli A., Oefner P.,
 RA Davis R.W.;
 RT "Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence."
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC002986; AAC17046.1; -
 DR InterPro; IPR003880; Pantane_attach.
 DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN_1.
 DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN_1.
 SQ SEQUENCE 204 AA; 22231 MW; 89592B28A691191 CRC64;

Query Match 3.4%; Score 7; DB 10; Length 204;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 170 KSALEA 176
 DB 122 KSALEA 128

RESULT 27
 O9JPN0 PRELIMINARY; PRT; 215 AA.
 AC O9JPN0;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, last annotation update)
 DE Putative lipoprotein GNA1162.
 GN GNA1162.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=487;
 OX
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=NGP165;
 RX MEDLINE=20175756; PubMed=10710308;
 RA Pizze M., Scariato V., Masignani V., Giuliani M.M., Arico' B.,
 RA Comanducci M., Jennings G.T., Baldi L., Bartolotti E., Capocchi B.,
 RA Galeotti C.L., Luzzi E., Manetti R., Marchetti E., Mora M., Nuti S.,
 RA Ratti G., Santini L., Savino S., Scarselli M., Storni E., Zuo P.,
 RA Broecker M., Hundt E., Knapp B., Blair E., Mason T., Tettelin H.,
 RA Hood D.W., Jeffries A.C., Saunders N.J., Granoff D.M., Venter C.,
 RA Moxon E.R., Grandi G., Rappelli R.;
 RT "Identification of Vaccine Candidates Against Serogroup B
 RT Meningococcus by Whole-Genome Sequencing.";
 RL Science 287:1816-1820(2000).

.....

NCBI TaxID=4896;

Schizosaccharomy

Schizosaccharomy

RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Devlin K., Churche C.M.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL:AL022103; CAA17889.1; -
 DR InterPro: IPR004366; Cyclin.
 DR Pfam: PF00134; cyclin; 1.
 DR SMART: SM00385; CYCLIN; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 229 AA; 25943 MW; EAA400A8F3906050 CRC64;

Query Match 3.4%; Score 7; DB 3; Length 229;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 IATSLD 46
 DB 109 IATSLD 115

RESULT 32

ID Q99SK1 PRELIMINARY; PRT; 229 AA.
 AC Q99SK1;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein SAV2051.
 GN SAV2051 OR SA1856.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OC Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Staphylococcus.
 NC NCBI_TaxID=158878, 158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
 RX MEDLINE=21311952; PubMed=11418146;

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiratsuku K.,
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 RT aureus".
 RL Lancet 357:1225-1240(2001).
 DR EMBL: AF003364; BAB58213.1; -
 DR EMBL: AF003364; BAB43138.1; -
 DR InterPro: IPR000905; Peptidase_M22.
 DR Pfam: PF00814; Peptidase_M22; 1.
 DR ProDom: PD002367; Peptidase_M22; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 229 AA; 25691 MW; 7F85E72BDFE8B61 CRC64;

Query Match 3.4%; Score 7; DB 16; Length 229;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 AKQQLDA 59
 DB 62 AKQQLDA 68

RESULT 33

ID Q8Y225 PRELIMINARY; PRT; 241 AA.
 AC Q8Y225;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Putative mannose-1-phosphate guanyltransferase-related protein
 DE (EC 2.7.7.-).
 GN RSC0511 OR RS04996.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 NC NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000.
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Brotlier P., Camus J.C., Cactolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Stigter P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weisendach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum".
 RL Nature 415:497-502(2002).
 DR EMBL: AL646059; CAD14039.1; -
 DR InterPro: IPR002224; 5_nucleotidase.
 DR InterPro: IPR001064; Cytostallin.
 DR InterPro: IPR002052; N6_Mtase.
 DR InterPro: IPR001825; NTP transferase.
 DR Pfam: PF00483; NTP transferase; 1.
 DR PROSITE: PS00785; 5_NUCLEOTIDASE_1; UNKNOWN_1.
 DR PROSITE: PS00225; CRYSTALLIN BETAGAMMA; UNKNOWN_1.
 DR PROSITE: PS00092; N6_MTASE; UNKNOWN_1.
 KW Transferase; Nucleotidyltransferase; Complete proteome.
 SQ SEQUENCE 241 AA; 25395 MW; 3B3518922CF41E4 CRC64;

Query Match 3.4%; Score 7; DB 16; Length 241;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 LDALISA 63
 DB 230 LDALISA 236

RESULT 34

ID Q8X4J5 PRELIMINARY; PRT; 249 AA.
 AC Q8X4J5;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein z1919.
 GN z1919.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NC NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirpatrick H.A.,
 RA Posfist G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamowski K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7".
 RL Nature 409:529-533(2001).
 DR EMBL: AE005333; AAG56008.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 249 AA; 24350 MW; E36010F654D6C77 CRC64;

Query Match 3.4%; Score 7; DB 16; Length 249;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 DAALSD 64
 Db 85 DAALSD 91

RESULT 35

Q92M58 PRELIMINARY; PRT; 262 AA.
 AC Q92M58;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Putative trifoliotoxin immunity protein.
 GN TXG OR RB0487 OR SMO2505;
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSymb (megaplasmid 2).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396508; PubMed=11481431;
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
 RA Golding B., Puelher A.,
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliloti."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 KM EMBL; AL603643; CAC4887.1; -
 SQ Plasmid; Hypothetical protein; Complete proteome.
 SO SEQUENCE 262 AA; 29321 MW; 359EC6ACFFD092F CRC64;

Query Match 3.4%; Score 7; DB 16; Length 262;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 SALLRAL 177
 Db 255 SALLRAL 261

RESULT 36

O25668 PRELIMINARY; PRT; 288 AA.
 AC O25668;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CO-chaperone-curved DNA binding protein A (CBPA).
 GN HPI024.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.,
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori."
 RL Nature 388:539-547(1997).
 DR EMBL; AE000610; AAD08066.1; -
 DR HSSP; P08622; 1BQZ.

DR TIGR: HPI024; -
 DR InterPro; IPR002939; DnaJ_C.
 DR InterPro; IPR001623; DnaJ_N.
 DR Pfam; PF00226; DnaJ; 1.
 DR Pfam; PF01556; DnaJ_C; 1.
 DR SMART; SM00271; DnaJ; 1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS50076; DnaJ_2; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 288 AA; 32908 MW; 914384C0768FFEE CRC64;

Query Match 3.4%; Score 7; DB 16; Length 288;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 DEIKKSY 190
 Db 18 DEIKKSY 24

RESULT 37

O92M23 PRELIMINARY; PRT; 288 AA.
 AC O92M23;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Putative CO-chaperone with DNAK.
 GN DnaJ 1 OR JHP0400.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.U., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
 RA Trust T.J.,
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori."
 RL Nature 397:176-180(1999).
 DR EMBL; AE001474; AAD05985.1; -
 DR HSSP; P08622; 1BQZ.
 DR InterPro; IPR002939; DnaJ_C.
 DR InterPro; IPR001623; DnaJ_N.
 DR InterPro; IPR003095; Hsp_DnaJ.
 DR Pfam; PF00226; DnaJ; 1.
 DR Pfam; PF01556; DnaJ_C; 1.
 DR PRINTS; PR00625; DnaJPROTEIN.
 DR SMART; SM00271; DnaJ; 1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS50076; DnaJ_2; 1.
 KM Complete proteome.
 SQ SEQUENCE 288 AA; 32863 MW; D31B6AC49862C472 CRC64;

Query Match 3.4%; Score 7; DB 16; Length 288;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 DEIKKSY 190
 Db 18 DEIKKSY 24

RESULT 38

O9KXY2 PRELIMINARY; PRT; 291 AA.
 AC O9KXY2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Hypothetical protein SC03866.
 GN SC03866 OR SCH18.03C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Brown S.P., Harris D.;
 RN Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RN Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleeser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleeser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,
 RA Huang C.-H., Kleeser T., Larke A., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Sanders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)." ;
 RL Nature 417:141-147(2002).
 DR EMBL; AL357152; CAB92986.1; -;
 DR InterPro; IPR000051; SAM_bind.
 KW Hypothetical protein.
 SQ SEQUENCE 291 AA; 31346 MW; 018C90DE1026952C CRC64;

Query Match 3.4%; Score 7; DB 16; Length 291;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 GRVATE 145
 Db 97 GRVATE 103

RESULT 39
 ID Q18362 PRELIMINARY; PRT; 292 AA.
 AC Q18362;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE C3A12.16 protein (Fragment).
 GN C3A12.16
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
 OC Rhabditidae; Peleodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilkinson J.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z68493; CA92799.1; -;
 FT NON TER 292
 SQ SEQUENCE 292 AA; 32823 MW; 5178A7AC1AB6A266 CRC64;

Query Match 3.4%; Score 7; DB 5; Length 292;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 SACOSTP 24
 Db 246 SACOSTP 252

RESULT 40
 ID Q9KSB6 PRELIMINARY; PRT; 292 AA.
 AC Q9KSB6;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE Hypothetical protein VC1311.
 GN VC1311.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Ginn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermlaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).
 DR EMBL; AE004211; AAF94469.1; -;
 DR TIGR; VC1311; -;
 DR InterPro; IPR000620; DUF6.
 DR Pfam; PF00892; DUF6; 2.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 292 AA; 31015 MW; BAB9FA93A42B028 CRC64;

Query Match 3.4%; Score 7; DB 16; Length 292;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 AAKSPK 163
 Db 282 AAKSPK 288

Search completed: July 8, 2003, 11:46:45
 Job time : 120 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 4, 2003, 03:13:42 ; Search time 45 Seconds
(without alignments)
1390.268 Million cell updates/sec

Title: US-10-069-544-2

Perfect score: 1043
Sequence: 1 MKIRVKPMVAMGLVLSAC.....EIKSYDKLSDYKLLSPDK 204

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_epool/US10069544/runat_30062003_091538_26300/app_query.fasta.1.391
-DB=Issued_Patents_NA -QFRT=fastcap -SUFFIX=emi -MINMATCH=0.1 -LOOCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pcio -NORF=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USFE=US10069544 @CGN 1.1.32 @runat_30062003_091538_26300 -NCPU=6 -ICPU3
-NO MMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -NARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

Database :

1: Issued Patents NA:*
2: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	143.5	13.8	5633	4	US-09-221-017B-545
2	106.5	10.2	693	4	US-09-134-001C-772
3	101	9.7	3476	2	US-08-649-046-1
4	90	8.6	3758	3	US-08-323-477-1
5	89.5	8.6	865	4	US-09-221-017B-606
6	89	8.3	13121	4	US-08-961-527-126
7	87	8.5	31880	4	US-09-453-702B-242
8	86	8.2	2145	4	US-09-059-584-48
9	86	8.2	2287	4	US-09-059-584-47
10	85.5	8.2	9223	4	US-08-961-527-59
11	85.5	8.2	46899	1	US-08-471-119A-1
12	84.5	8.1	993	4	US-09-134-001C-574

13	83.5	8.0	2121	4	US-09-059-584-45	Sequence 45, Appl
14	83.5	8.0	3957	1	US-07-689-008-5	Sequence 5, Appl
15	83.5	8.0	9540	1	US-07-689-008-1	Sequence 1, Appl
16	83.5	8.0	30549	4	US-09-134-001C-322	Sequence 322, Appl
17	82	7.9	2217	1	US-07-618-946B-24	Sequence 24, Appl
18	82	7.9	2442	1	US-07-618-946B-25	Sequence 25, Appl
19	82	7.9	4054	4	US-07-618-946B-2	Sequence 2, Appl
20	82	7.9	43360	4	US-09-453-702B-206	Sequence 206, Appl
21	82	7.9	43325	4	US-09-453-702B-261	Sequence 261, Appl
22	80.5	7.7	609	4	US-09-328-111-842	Sequence 842, Appl
23	80.5	7.7	8931	3	US-09-051-019-1	Sequence 1, Appl
24	80.5	7.7	16836	4	US-09-147-236-10	Sequence 10, Appl
25	80.5	7.7	16836	4	US-09-147-236-10	Sequence 10, Appl
26	79.5	7.6	8821	4	US-09-221-017B-1028	Sequence 1028, Appl
27	79.5	7.6	1985	1	US-07-792-865D-2	Sequence 2, Appl
28	79	7.6	830	4	US-09-221-017B-1087	Sequence 1087, Appl
29	78.5	7.5	581	4	US-09-221-017B-607	Sequence 607, Appl
30	78.5	7.5	1332	2	US-08-795-475-5	Sequence 5, Appl
31	78.5	7.5	1446	2	US-08-569-150A-2	Sequence 2, Appl
32	78.5	7.5	2868	3	US-08-274-121B-1	Sequence 1, Appl
33	78	7.5	4700	2	US-08-928-692-16	Sequence 16, Appl
34	78	7.5	4700	4	US-09-339-972-16	Sequence 16, Appl
35	77.5	7.4	1310	4	US-09-047-288-1	Sequence 1, Appl
36	77.5	7.4	1310	4	US-08-802-191-1	Sequence 1, Appl
37	77.5	7.4	2180	3	US-09-286-904-23	Sequence 23, Appl
38	77.5	7.4	2180	4	US-09-640-101-23	Sequence 23, Appl
39	77	7.4	1248	4	US-09-134-001C-644	Sequence 644, Appl
40	77	7.4	5077	2	US-08-687-956A-22	Sequence 22, Appl
41	77	7.4	18475	4	US-08-961-527-38	Sequence 38, Appl
42	76.5	7.3	1502	2	US-08-651-940-1	Sequence 1, Appl
43	76.5	7.3	1502	4	US-09-295-029-1	Sequence 1, Appl
44	76.5	7.3	2721	4	US-09-221-017B-965	Sequence 965, Appl
45	76	7.3	2067	2	US-08-713-928B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-09-221-017B-545
Sequence 545, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P11182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:
NAME: Montoy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 545:
SEQUENCE CHARACTERISTICS:
LENGTH: 5633 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORGANISM: FORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1...5633
US-09-221-017B-545

Alignment Scores:
Pred. No.: 8,04e-09 Length: 5633
Score: 143.50 Matches: 54
Percent Similarity: 45.45% Conservative: 31
Best Local Similarity: 28.88% Mismatches: 55
Query Match: 13.76% Indels: 47
DB: 4 Gaps: 10

US-10-069-544-2 (1-204) x US-09-221-017B-545 (1-5633)
QY 37 ArgThGlnIleAlaIleSerLeuAspMetGlyLeuAsp-----GlnAlaIys 54
DB 1994 CAAGCACATATAGCTCTCAGCGGATAGAGCTGGGAGAGAGATACCGTCGAGCATT 2053
QY 55 GlnGlnLeuAspAlaIleAspSerAlaAspArgGlnPheAlaProAlaIleTyrArgGlnLeu 74
DB 2054 CGGAACTGAAACGAGGACTTCGATGAGCTCTCTTCCTCCTCGCTATGCGAGATG 2113
QY 75 AlaIysValIleTyrGlnAlaSerGlnAspAlaThrHisGlnThrIleTyrAspIleArgLeuPhe 94
DB 2114 GCATGCTGCACCTCAAGAGCAG-----CGCAAGCAGAGGCTATGAGACCCCTG 2164
QY 95 GlnIleValIleGlnLeuAspProIleAspMetGlnSerTyrMetAspTyrGlyPheTyr 114
DB 2165 GACAAAGCATATGAAATTTGGAGCCGAGCACTGTCCAACTATATCAACCGGGGTGCATC 2224
QY 115 LeuValGlnMetGlyAspLeuSerGlyAlaLeu-----IleTyrPheAsp 129
DB 2225 CGTACCAATCGAAGCACTTCGCGAGCAGCACTACAGCAAGCAGTTCGCCAA 2284
QY 130 LysPro-----SerArgAlaIle-----GlyTyrGlnGlyArg 140
DB 2285 AAGCCGAACGACAACTGGACGTTTCAACCGTCTTCCTCCGCTTACCTCGGAGC 2344
QY 141 Val-----ValAlaIleGlnAspMetAlaTyrIleTyr-----TyrHis 153
DB 2345 GTGAACATATGATCGAGGATTCGATGATGATTCGCTGAGCGGAGCAATATCAT 2404
QY 154 GlnTyrGlnAlaIleAspSerProThrIleAspAspTyrAspAspAlaIle 173
DB 2405 GCACCTC-----TACAAACCGTGC----- 2422
QY 174 GlnArgAlaIleIleSerGlyThrGlnHisAspGlnIleLysSer-----Tyr 190
DB 2423 -----ATCTCTCAACGAGTCAACAGAAACGAAAGCATTACGCACTTC 2470
QY 191 AspLysLeuLeuSerAspTyr 197
DB 2471 GACAAAGTGTGGGACACTAT 2491

RESULT 2
US-09-134-001C-772
Sequence 772, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 772
LENGTH: 693
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-772

Alignment Scores:
Pred. No.: 3.41e-05 Length: 693
Score: 106.50 Matches: 36
Percent Similarity: 46.56% Conservative: 25
Best Local Similarity: 27.48% Mismatches: 53
Query Match: 10.21% Indels: 17
DB: 4 Gaps: 5

US-10-069-544-2 (1-204) x US-09-134-001C-772 (1-693)
QY 75 AlaIysValIleTyrGlnAlaSerGlnAspAlaThrHisGlnThrIleTyrAspIleArgLeuPhe 94
DB 37 AGTAAGATATACCAAAATATTAAAGATGTAAATAGTACCGCTTGAAGACTTGT 96
QY 95 GlnIleValIleGlnLeuAspProIleAspMetGlnSerTyrMetAspTyrGlyPheTyr 114
DB 97 GAAAT-----ATGAAAGAAATCCGCCATTTGTAATATTAATTAACGAGTATGTA 153
QY 115 LeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspIleProSerArgAla 134
DB 154 TTATCAGATGTTGGAAATTTGAAAGCAGACGTTTTTCCAAAGCATTAACTATT 213
QY 135 IleGlyTyrGlnGlyArgValAlaIleGlnAspMetAlaTyrIleTyrTyrHisGln 154
DB 214 GAACCTGAAATGTGCTGT-----TATTATACCTTGGCAATATTATTATATGAA 267
QY 155 TyrGlnAlaIleAspSerProThrIleAspAspTyrAspAspAlaIleGln 174
DB 268 -----GACGTTTAAAGCAAGCAGTAAAGCTATATATCA 300
QY 175 ArgAlaLeuIleSerGlyThrGlnHisAspGlnIleLysSerTyrAspIleLeu 194
DB 301 ACTGCACCTA-----CAATATGATAGAGCAAAAGATTGTAATATATGAT 348
QY 195 Ser-----AspTyrIleLeuLeuSerAspTyrIle 204
DB 349 GGTATGCTTTATATCACTAGTCTTTTAA 381
RESULT 3
US-08-649-046-1
Sequence 1, Application US/08649046
Patent No. 5912415
GENERAL INFORMATION:
APPLICANT: OLSZEWSKI, NEIL E.
APPLICANT: JACOBSEN, STEVEN E.
TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF
TITLE OF INVENTION: IDENTIFICATION AND USE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: MÖETTING, RAASCH, GEHARDT & SCHWABPACH, P.A.

STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649, 046
FILING DATE: 16-MAY-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCCORMACK, MYRA H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110,00340101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1225
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3476 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-649-046-1

Alignment Scores:
Pred. No.: 0.00234 Length: 3476
Score: 101.00 Matches: 36
Percent Similarity: 39.19% Conservative: 22
Best Local Similarity: 24.32% Mismatches: 64
Query Match: 9.68% Indels: 26
DB: 2 Gaps: 4

US-10-069-544-2 (1-204) x US-08-649-046-1 (1-3476)

QY 33 LeuAlaGlnIleValGThGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGln 52
DB 801 TTAGCATTGTTTGGACCGACCTTGAAGCTGCGCTGGAATATCTGACGAA 860
QY 53 AlaLysGlnIleuAspAlaIleuSerLysAspGlnPheAlaProAlaTyrArg 72
DB 861 GGAATTCAAAAGTATTACGAAGCCCTTAAAGATTGACCACTATGCTCTGCAATATTAC 920
QY 73 ThrLeuAlaLysValTyrGlnIleAspSerGlnuAspAlaThrHisGlnThrLysAlaGlnArg 92
DB 921 AACTTAGGTGTTGTATAC-----TCCGAATGATGCATATGACAAATGCCCTTGAGC 971
QY 93 LeuPheGlnLysValAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112
DB 972 TGCTACGAAGAGCGCTGACCTGAGAGCCCTATGTCATGTCGAAGCATATTGTAACATGGGT 1031
QY 113 PheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLys----- 130
DB 1032 GTCATTATTAAGAACCGGTGCTGATCGAATGCAATCTGTTATGAGAGATGCTTA 1091
QY 131 -----ProSerArgAlaIleGlyTyrGlnGlyAspValAlaIle----- 144
DB 1092 GCTGTGCTCCAAACTTGGAGTTGCGAAGAACATATATGGCATGCTCTGACAGATTTA 1151
QY 145 -----GluAsnMetAlaTyr----- 149
DB 1152 GGAACAAAGCTTAAACTTGAAGCGATGTAAACCAAGAGTGGCATATTACAAAGAGCT 1211
QY 150 IleTyrTyrHisGlnTyrGlnAla 157
DB 1212 CTTATTTAATCTGGCATATGCA 1235

RESULT 4

US-08-323-477-1
Sequence 1, Application US/08323477
Patent No. 6086896

GENERAL INFORMATION:
APPLICANT: Sparling, P. Frederick
TITLE OF INVENTION: ANTIGENIC IRON REPRESSIBLE PROTEINS FROM
TITLE OF INVENTION: N. MENINGITIDIS RELATED TO THE HEMOLYSIN FAMILY OF TOXINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Imclone Systems Incorporated
STREET: 180 Varick Street
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10014

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,477
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER: US/07/920,963

FILING DATE: 28-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Felt, Irving N.

REGISTRATION NUMBER: 28,601

REFERENCE/DOCKET NUMBER: SPA-2-2P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-645-1405

TELEFAX: 212-645-2054

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3758 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORGANISM: Neisseria meningitidis

FEATURE:

NAME/KEY: CDS

LOCATION: 298..3645

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 298..3642

US-08-323-477-1

Alignment Scores:
Pred. No.: 0.0833 Length: 3758
Score: 90.00 Matches: 58
Percent Similarity: 36.40% Conservative: 33
Best Local Similarity: 23.20% Mismatches: 69
Query Match: 8.63% Indels: 90
DB: 3 Gaps: 15

US-10-069-544-2 (1-204) x US-08-323-477-1 (1-3758)

QY 1 MetLysIleArgValLysTyrProMetValMetAlaMetGlyLeuValLeuSerAlaCys 20
DB 1951 ATGCACTTCAACCGATTGACGCAACGCGTATGAGGTATGCACTG----- 2001
QY 21 GlnSerThrProIleProIleProLysAsnAsnProGlnIleuAlaGlnIleValGThGlnIle 40
DB 2002 -----ACACCA-----TCCCAAGTACGACAACTAAAAAGAACGCT 2037

Qy	41	AlalleSerLeuAspMetGlyLys-----LeuAspGlnAlaLysGln-----	56
Db	2038	TTAGTTCCTCCCTTTCTGTAATAAGCTTAAGACAGCTATTGACGCCGCCGAGCAGCTTCC	2097
Qy	57	---LeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAla	75
Db	2098	GTGCTTGTATGCTCTACACGGGCGAGATTTCACG-----ACACTC---	2136
Qy	76	LysValTyrGlnAlaSerGluAspAla-----ThrHis	86
Db	2137	---TATTACATGAGGAGAAAGACGCCGCTTAATATCTGTCAAGTACCAACGATCATATC	2199
Qy	87	GlnThrLysAlaGlnArgLeuPheGlnLysAlaIle-----GluLeuAsnProLys	103
Db	2194	GACCATCTCCGCCAATAACATCTACCAAAACCTGTTGCCAAACCCGTTTGAGCCA---	2250
Qy	104	AspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGly-----	119
Db	2251	---TATTGTAATCAAAATATGTTCCAAATGGAA	2280
Qy	120	---AspLeuSerGlnAlaLeuIleTyrPheAsp-----	129
Db	2281	AATGATACGTTCACTTGTGATTTTGTGTCTTGTTCACAGCATTTAACCATGTCAAAGA	2340
Qy	130	---LysProSerArgAlaIle-----	135
Db	2341	ACTAATCCGCAAAAAGCTTTTGTGATTTGGCCGAGATGCTTCATATGGCGAATCTGT	2400
Qy	136	---GlyTyrGlnGlyArgValValAlaIleGlnAsnMetAlaTyrIleTyrThrGln	154
Db	2401	TCTTGTAATGAAGGCCGGAAGACTTAATGCCGCAAT-----TATGTG	2433
Qy	155	TyrGlnAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGln	174
Db	2440	GAGGAGGCAAAAAAGCAGGTAATTTGAAGATTACAGAAAGTGTGGGTGACGAGACAC	2493
Qy	175	ArgAlaLeuIle-----SerGlyThrGlnHisAspGlnIleLysLysSer-----	189
Db	2500	GTTGCATTATTATGCTAAACATCGGGTACGCAAGCAGATGATATCTCGCAAAATGTAGGC	2555
Qy	190	---TyrAspLysLeuLeuSerAspTyr	197
Db	2560	TTTGTCTAATATATTAATGTTCTTTATAT	2589
RESULT 5			
US-09-221-017B-606/c			
: Sequence 606, Application US/09221017B			
: Patent No. 6444799			
GENERAL INFORMATION:			
APPLICANT: Ross, Bruce C.			
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF			
NUMBER OF SEQUENCES: 1120			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: MORRISON & FOERSTER			
STREET: 755 PAGE MILL ROAD			
CITY: Palo Alto			
STATE: CA			
COUNTRY: USA			
ZIP: 94304-1018			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Diskette			
COMPUTER: IBM Compatible			
OPERATING SYSTEM: Windows			
SOFTWARE: PasteSeq for Windows Version 2.0b			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/221, 017B			
CLASSIFICATION:			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: P1182			
FILING DATE: 31-DEC-1997			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: P11546			

1 FILING DATE: 30-JAN-1998
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: PP2911
4 FILING DATE: 09-APR-1998
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: PCT/AU98/01023
7 FILING DATE: 10-DEC-1998
8 ATTORNEY/AGENT INFORMATION:
9 NAME: Montroy, Gladys H
10 REGISTRATION NUMBER: 32,430
11 REFERENCE/DOCKET NUMBER: 27340-20021.000
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: 650-813-5600
14 TELEFAX: 650-494-0792
15 TELEX: 706141
16 INFORMATION FOR SEQ ID NO: 606:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 865 base pairs
19 TYPE: nucleic acid
20 STRANDEDNESS: double
21 TOPOLOGY: circular
22 MOLECULE TYPE: DNA (genomic)
23 HYDROTICAL: NO
24 ANTI-SENSE: UNKNOWN
25 ORIGINAL SOURCE:
26 ORGANISM: PORPHYROMONAS GINGIVALIS
27 FEATURE:
28 NAME/KEY: misc feature
29 LOCATION: 1...865
30 OS-09-221-017B-606

Alignment Scores:

Pred. No.:	0.0096	Length:	86
Score:	89.50	Matches:	40
Percent Similarity:	37.87%	Conservative:	24
Best Local Similarity:	23.67%	Mismatches:	50
Query Match:	8.58%	Indels:	5
DB:	4	Gaps:	7

US-10-069-544-2 (1-204) X US-09-221-017B-606 (1-865)

```

Oy      40  TlAlAlIleSerLeuLeuAspMetClyLyLeuAspGlnAlaIySgInGlnLeuAspAla 59
Db      817  ATAGGAGITTCGATATTGTGGAGAAAGCATCGTATGAGAAGCGATTAAAGACTATTCTTAAG 758
Oy      60  AlAlenSerAlaAspArgGlnPheAlaProAlaTyraGlnThrLeuAlaIySvalTyrgln 79
Db      757  GCATAGCAATTAGACGATTAATTAACCCCTCGTATCATGTTAGAGAAATGCATAT--- 701
Oy      80  AlASeGluAspAlaThrHisGlnThrIySalaglnArgLeuPheGluPheAlaIleGlu 99
Db      700  ---TCTAAGAAAGCATGTATATAA---AAAGCGATTAAAGCTATTCTCAGGCAATAGAA 647
Oy      100  LeuAsnProLyAspMetGlnSerTyI--- 108
Db      646  TTAGACGATTAATTCGCCCATCCGATATATGTTAGAGAAATGCATATTGTGAGAAAGGA 587
Oy      109  ---MetAspTyI--- 111
Db      586  TCATATGAGAAGACGATTAAAGACTATTCTCAGGCATATGATTAAGCATTAATAACGCC 527
Oy      112  ---GlyPheTyILeuValGlnMet 118
Db      526  CCTGCGTATCATCGTATGAGAGAAATGCATATTTCTAAGAAAGAGATGGTATTAATAAAGCATT 466
Oy      119  GlYAspLeuSerGlyAlaLeuIleTyIrpheaSplysProSerArgAlaIleGlyTyrglu 138
Db      466  AAGACCTATTCTCAGGCATATGAAATTAAGCATATAA---TACATCTCTGGGTATATAT 413
Oy      139  GlYArgValValaAlaIleGluAsnMetChalTyIleTyIrpHisGlnTyGlnAlaIala 158
Db      412  GGTAGAGAGATGCA---TATATATGAGATATGAGAGATTTTGAA 374

```

QY 159 LysSerProThrLysAspAspTyrAsn 167
DB 373 AAAAGT-----CAAGCGACTTCAAT 353

RESULT 6
US-08-961-527-126/C
Sequence 126, Application US/08961527
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 126:
SEQUENCE CHARACTERISTICS:
LENGTH: 1121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-08-961-527-126

Alignment Scores:
Pred. No.: 0.794 Length: 13121
Score: 89.00 Matches: 36
Percent Similarity: 41.36% Conservative: 31
Best Local Similarity: 22.22% Mismatches: 73
Query Match: 8.53% Indels: 22
DB: 4 Gaps: 4

US-10-069-544-2 (1-204) x US-08-961-527-126 (1-13121)

QY 22 SerThrProIleProProLysAspAsnProGlnLeuAlaGlnIleArgThrGlnIleAla 41
DB 9831 AGTCCTCAGCTTCTCAAAAACAGACATCACTAAGAAAGCTTAAAGAGCAATTAAGT 9772

QY 42 IleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaIleu 61
DB 9771 ATAGCT-----AAAGAGAGGCTAGTCAAGAAAGAGAGCTTGAATGCCAAG 9721

QY 62 SerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSer 81
DB 9720 GAGCAGCTTCAGAAATTAATGGCTGACTACCAAGCTATGCCAAG---GAGCAAGAGAG 9664

QY 82 GluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsn 101
DB 9663 CAGAAATACTCTCTACAGCTCAACAAAGCCCAACTCTTTGACCGCTGATTAATCTCAA 9604

QY 102 ProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeu 121

DB 9603 AACAGAGCGCCAGAGCTCAAAAGTTGGAAATATCTGAGAAATCATAGTAATTTAT 9544

QY 122 SerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGlnGlyArgVal 141
DB 9543 GCGAGTGTTAAGAGCTGTTCTTCCAAAGAAAAGATGCGCTAGGTGGATTAATTTGCGACTC 9484

QY 142 ValAlaIleGluAsnMetAlaTyrIleTyrTyrHisGlnTyrGlnAlaAlaLysSerPro 161
DB 9483 -----ACTGCGAATCTGACCTTGATGTTATATCA----- 9451

QY 162 ThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGlyThr 181
DB 9450 -----ACTGCGCTAGAGATTCCTTAGGGGCAAGTAGC 9418

QY 182 GlnHis 183
DB 9417 CAGCAT 9412

RESULT 7
US-09-453-702B-242/C
Sequence 242, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plinkney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27986
REFERENCE/DOCKET NUMBER: 960296,95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 242:
SEQUENCE CHARACTERISTICS:
LENGTH: 31880
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 242:
US-09-453-702B-242

Alignment Scores:
Pred. No.: 5.9 Length: 31880
Score: 87.00 Matches: 49
Percent Similarity: 38.03% Conservative: 40
Best Local Similarity: 20.94% Mismatches: 77
Query Match: 8.34% Indels: 68

Qy 145 GluAsnMetAlaTyrIleTyrThiIleGlnTyrGluAlaAlaIleSer---ProThrIys 163
Db 772 -----GGCGGTGTTTATATATGCTCAACGACCGCCAAAGACGTGCCACACAA 822
Qy 164 Asp-----AspTyrAsnAsnAlaIleSerAlaLeuGluAlaLeuIleSerGlyThr 181
Db 823 GATGGGTCAATATATAAGACATTGGACCTTATGACCGATGTTGCCAAAAAAGAAAC 882
Qy 182 GlnHisAspGluIleIleValSerTyr 190
Db 883 CGATTACCGAAGTAAAGAAACCTAT 909

RESULT 9
US-09-059-584-47
Sequence 47, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Harkness, Robin E
APPLICANT: Loomore, Sheena W.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 2287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-059-584-47

Alignment Scores:
Pred. No.: 0.135 Length: 2287
Score: 86.00 Matches: 51
Percent Similarity: 35.81% Conservative: 31
Best Local Similarity: 22.27% Mismatches: 73
Query Match: 8.25% Indels: 74
DB: 4 Gaps: 13

US-10-069-544-2 (1-204) x US-09-059-584-47 (1-2287)
Qy 11 MetaIaMetGlyLeuValIleuSerAlaCys-----GlnSerThrProIle 25

Db 440 ATGGGTATGCGACATGGCTTTGAGTAAATTAATCTATACGACCAACAGACCGCATTA 499
Qy 26 ProIleIysAsnAsnProGlnIleuAlaGlnIleThrGlnIleAlaIleSerIleu 45
Db 500 GATCCAAAAAT-----ATCATTAACCTTA 523
Qy 46 AspMetGlyIys-----LeuAspGlnAlaIleGlnIleuAspAlaAlaIleuSerAla 63
Db 524 GACCGTAAAAACAAGTTGCTGACATCAAAATCACCATTTGCCATTTTGGTTGATGTA 583
Qy 64 AspArgGlnPheAlaProAlaTyrArgThrIleuAlaIleValTyrGlnAlaSerGluAsp 83
Db 584 GAAATTAATTCCTTGATGCTAT-----ATACCAAAAATGATGAAGCGGATTAATAAT 637
Qy 84 Ala-----ThriIleGlnThrIys----- 89
Db 638 GCCATTGGTGAAGAATTAAGAGAAAATGACAAAATTAATAAATATCCGATGAGAA 697
Qy 90 ---AlaGlnArgLeuPheGlu----- 95
Db 698 CTTCGCAAAAATCAAGAAATGTGGTAAAGCCCTGAGTTTCAGCAAGTATTATCA 757
Qy 96 -----LysAlaIleGluLeuAsnProIleAsp 104
Db 758 TCGATTAAGCGAAACCTTCATTCAATGACAAACAAACCAAGCAACCAACAGAT 817
Qy 105 MetGlnSerTyrMetAspTyrGlyPheTyrIleuValGlnMetGlyAspLeuSerGlyAla 124
Db 818 TTAATAA---TATGTTGATTAATGTTACTACTTGGTG-----AATGATGCCAATTATCTA 868
Qy 125 LeuIleTyrPheAspIysProSerArgAlaIleGlyTyrGlnGlyArgValAlaIle 144
Db 869 ACCGTCAAAACAGCAAAACCAAA-----CTTGGAATTGAGTCCGTG----- 913
Qy 145 GluAsnMetAlaTyrIleTyrThiIleGlnTyrGluAlaAlaIleSer---ProThrIys 163
Db 914 -----GGCGGTGTTTATATATGCTCAACGACCGCCAAAGACGTGCCACACAA 964
Qy 164 Asp-----AspTyrAsnAsnAlaIleSerAlaLeuGluAlaLeuIleSerGlyThr 181
Db 965 GATGGGTCAATATATAAGACATTGGACCTTATGACCGATGTTGCCAAAAAAGAAAC 1024
Qy 182 GlnHisAspGluIleIleValSerTyr 190
Db 1025 CGATTACCGAAGTAAAGAAACCTAT 1051

RESULT 10
US-08-961-527-59
Sequence 59, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-961-527-59
;
Alignment Scores:
Pred. No.: 1.38 Length: 9223
Score: 85.50 Matches: 34
Percent Similarity: 44.53% Conservative: 27
Best Local Similarity: 24.82% Mismatches: 53
Query Match: 8.20% Indels: 23
Gaps: 7
;
US-10-069-544-2 (1-204) x US-08-961-527-59 (1-9223)
;
QY 28 LysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuAspMet 47
;
DB 8463 AAAAATGAAAAAGATTGGCAAGCTCCGAAAGCTGATTTCTTCATCTTATCAAGAA 8522
;
QY 48 GlyLeuAspGlnAlaIleGlnGlnIleuAspAlaIleuSerAlaAspArgGlnPhe 67
;
DB 8523 GAAAGCTTATCCAAAGAAAAAGAA-----CTTCCAGCAATGAGAACTTC 8567
;
QY 68 AlaProAlaIleArgThrLeuAlaIleGlnAlaIleSerGlnAspAlaThrIleGln 87
;
DB 8568 TGGAGCTCCTATCAAGAG-----TTCAAGAGTAAACAGAGATGCCGTTTATGAA 8615
;
QY 88 ThrLysAlaGlnArgLeuPheGlnIleGlnIleuAsnProLysAspMetGlnSer 107
;
DB 8616 -----TTTGAG-----GTGGAGTTTCACTCAATCAAAATGAAAAA 8651
;
QY 108 TyrMetAspTyrGlyPheTyrLeu-----ValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
;
DB 8652 GATGTCGATGATGATTTTCGTCAGAGTCAGCTGTTGATTTGTCAGAGAGGACTTATC 8711
;
QY 127 TyrPheAspLysProSerArgAlaIleGlyTyrGlnGlyArgValAlaIle----- 144
;
DB 8712 TTTGTG-----CCGAACATGACGCTTGATGATGAGAGATGAAGTGAAGTTTCATC 8765
;
QY 145 ----GluAsnMetAlaTyrIleTyrTyrIleGlnIleGlnAlaIleAlaIleSer 160
;
DB 8766 AGGGAACCAAGCTCCTATGATGTCACCAAAAGACGCTGCCGAGAAAAAT 8816
;
RESULT 11
US-08-471-119A-1
; Sequence 1, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergerdorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
;
US-08-471-119A-1
;
Alignment Scores:
Pred. No.: 17.2 Length: 46899
Score: 85.50 Matches: 34
Percent Similarity: 44.53% Conservative: 23
Best Local Similarity: 26.56% Mismatches: 48
Query Match: 8.20% Indels: 23
Gaps: 6
;
US-10-069-544-2 (1-204) x US-08-471-119A-1 (1-46899)
;
QY 23 ThrProIleProProLysAsnAsnProGlnLeuAlaGlnIle-----Arg 37
;
DB 13983 ACACCAATGCGCCAAATCGCAATGCAAGCTGAGCAATATGCGCTTTCGTCAAT 14042
;
QY 38 ThrGln-----IleAlaIleSerLeuAspMetGlyLysLeuAspGlnAlaIleGlnGln 56
;
DB 14043 ACGCATGATATACGATTAACCTCAACAGAGACGACCTTTGATCTACTAGTCAGCAG 14102
;
QY 57 LeuAspAlaAlaLeuSerAla-----AspArgGlnPheAlaProAlaTyrArgThrLeu 74
;
DB 14103 GTCGGTCAACGGCGACAGCTGCATTCGCCATCAGACGTCCTCGTTCGAGAAATCGTC 14162
;
QY 75 AlaLysValIleGlnAlaIleSerGlnAspAlaThrIleGlnIleThrLysAlaGlnArgLeuPhe 94
;
DB 14163 TCTACTCTTTGCGCGGTTTCGAGATGATCCGAAACCACTTGTCAGCTCATGTTT 14222
;
QY 95 -----GluLysAla-----IleGlnLeu 100
;
DB 14223 GCGGTGCAATTCGCAAGAACTCGGTGAGCTGAAAGCTGAGAAACGCTCACAGCAGGT 14282
;
QY 101 AsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAsp 120
;
DB 14283 GTTCCACAGAGATACAGACCCGGTTCGATTTGGAATTCACCTCTTCCAGCAAGATGAC 14342
;
QY 121 ----LeuSerGlyAlaLeuIleTyr 127
;
DB 14343 AAGCTTGAGGCTCATCTCTAT 14366
;
RESULT 12
US-09-134-001C-974
; Sequence 974, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

```

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 PRIOR FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 974
 LENGTH: 993
 TYPE: DNA
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-974

Alignment Scores:
 Pred. No.: 0.0592 Length: 993
 Score: 84.50 Matches: 56
 Percent Similarity: 31.79% Conservative: 40
 Best Local Similarity: 18.54% Mismatches: 75
 Query Match: 8.10% Indels: 131
 DB: 4 Gaps: 13

US-10-069-544-2 (1-204) x US-09-134-001C-974 (1-993)

Qy 10 ValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProLeuProPheProLysAsn 29
 46 GTAACAGCAAGTGCATTTGTTAGCTCTGCTGTTCAATGCACTGAATCGAAGAT 105
 Qy 30 AenProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSer-----LeuLeuAspMet 47
 106 AACACACTATTCTTCTAAAGCTGCTGATGTTAAAGTTCGACGTAAAGAAAAATG 165
 Qy 48 GlyLys----- 49
 166 GGCAAGAACAAATTGCTAACACATCTTTAGCATTTGATTAAACAAAGTTTACAGAT 225
 Qy 49 ----- 49
 226 AAATATTAAGATTAAGTATGATCAAAAGATTTGTAAGATTAATAAGAAAAAGAGAAA 285
 Qy 50 -----LeuAsp 51
 286 CAATACGGCGGTAAGATCAATTTGAAGCATGTTAAACAAAGATATGCACTTAT 345
 Qy 52 GlnAlaLysGlnGln-----LeuAspAlaAla 60
 346 GATTATTAAGAACAGAAATAATTATCAGCTTACAGAAACAATTGTTACTAGATTAAGTT 405
 Qy 61 LeuSerAlaAspArgGlnPhe-----AlaProIaIyArgThrLeuAlaLys 76
 406 AATGATACAGATTAAGAAATCAAGAAATCTTAAAAACATCAACATTTTAATTAA 465
 Qy 77 ValTyrGlnAlaSer-----GluAspAlaThrIleGlnThrLysAla 90
 466 GTTAAATCAATATCAGGATTAAGAGATTATACAGATTAATAAGGCAAGAAAGAAACT 525
 Qy 91 GlnAlaGluPheGluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAsp 110
 526 GAAAAAATTT-----CAAAAAAGAGTGAAGAAAGATCCAAAT----- 561
 Qy 111 TyrGlyPheTyrLeuValGlnMetGlyAspLeuSer----- 122
 562 -----AAGTTTGAGAAATATAGCAAAAAAGAAATCAATGACACGTTCT 603
 Qy 123 -----GlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyr 137
 604 TCAGCTAAGAAAGATGTTACATTAGTATGATTAAGGCAATGATGATAGCTTC 663
 Qy 138 -----GluGlyArgValIleAlaIleGlnAsnMetAlaTyr 149
 664 GAAAAAGCACTATTAAATTAAAGAGAGAGAGTTCTTAAGTGTAAAGAAACAGACTAT 723

Qy 150 IleTyrThrIleGlnTyrGluAlaAlaLysSerProThrLysAspPheTyrAsnAla 169
 724 -----GGCTATCATATTATTAAGACATTAAGAAATCTGATTTTAATAGGAA 771
 Qy 170 LysSerAlaLeuGlnArgAlaLeuIleSerGlyThrGlnIleAspGluIleLysLys 188
 772 AAATCAATATCAAAACAAATTAATC-----GAGAAAGGTACAGAGAAA 819
 Qy 189 -----SerTyrAspLysLeuLeuSerAspTyrLysLeuLeuSerAsp 202
 820 CTTAAATTATTACTGATGATCAAGAAATTTAAAGAGTAAAGTA-----GAC 873
 Qy 203 TyrLys 204
 874 TACAAA 879

RESULT 13
 US-09-059-584-45
 Sequence 45, Application US/09059584
 Patent No. 6440701
 GENERAL INFORMATION:
 APPLICANT: Myers, Lisa E
 APPLICANT: Schryvers, Anthony B
 APPLICANT: Harkness, Robin E
 APPLICANT: Loosmore, Sheena M.
 APPLICANT: Du, Run-Pan
 APPLICANT: Yang, Yan-Ping
 APPLICANT: Klein, Michel H
 TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
 NUMBER OF SEQUENCES: 60
 CORRESPONDENCE ADDRESS:
 STREET: 6th Floor, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/059,584
 FILING DATE: 14-APR-1998
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/778,570
 FILING DATE: 03-JAN-1997
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Stewart, Michael I
 REGISTRATION NUMBER: 24973
 REFERENCE/DOCKET NUMBER: 1038-794
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 45:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2121 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-059-584-45

Alignment Scores:
 Pred. No.: 0.263 Length: 2121
 Score: 83.50 Matches: 52
 Percent Similarity: 36.82% Conservative: 36
 Best Local Similarity: 21.76% Mismatches: 84
 Query Match: 8.01% Indels: 67
 DB: 4 Gaps: 13

US-10-069-544-2 (1-204) x US-09-059-584-45 (1-2121)

```

QY      11 MetAlaMetGlyLeuValLeuSerAlaCys-----GlnSerThrProIle 25
      298 ATGGCTTATGGCATGCGCTTGTAGTAAATGAATCTACACAAACAAACAGACGCGCATTA 357
      26 ProProlAsn-----AsnProGlnLeuAlaGln----- 35
      368 GATGAAAAAGATATCTATTACTTAGACGCTAATAAACAAGTTCGAAAGAGTAAATATCG 417
      36 -----IleArgThrGlnIleAlaIleSerLeuLeuAsp-----MetGlyLysIleu 50
      418 CCATTGCCATTTTCGTTGGATGTAGAAAAATAATTCCTTGAAGCTTATATGCAAAAAATG 477
      51 AspGlnAla-----LysGlnGlnLeuAspAlaAla 60
      478 AATGAAGCGGATMAAAATGCCATTGTCACAGATTAAAGAAATTAATAAGCAAGTCA 537
      61 LeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAla 80
      538 TTATCTAA-----GCAGAGCTTGCCTAA-----CAATTC 567
      81 SerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPhe----- 94
      568 AAGAAGATGTCGCTAAAGCCATGAGTTACAGCAAGTATATCATCTACTGAAAAACAA 627
      95 -----GluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMet 109
      628 ATTTTCATTCATCAATATGAGACACACCAACCAACCAACGAGATTACAA---TATGTT 684
      110 AspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAsp 129
      685 GATTATGCTTACTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
      130 Lys-----ProSerArgAlaIleGlyTyrGlnGlyArgValAlaIle 144
      739 GAACCTTGAAATTAGGCGCTGCGCGCTGTGTTTATATGTCACCAACGACCGCCAA 798
      145 Glu-----AsnMetAlaTyrIleTyrThrIleGlnTyrGlnAlaAlaLys 159
      799 GACCTACCCACACACAGATGCGCTCAATATTAAGACATGGAGCTTATGACCGATGTT 858
      160 SerProThrLysAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSer 179
      859 GCCAAACAAAGAAACCGATTAGCGAAGTGAAGAAACCTTCACAGCAGCTCGTATTAT 918
      180 GlyThrGlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSerAspTyrLys 198
      919 GGAACATCTTCA-----AAGATGAATACACCGCTTATTAAGTATGAGAA 966

```

RESULT 14

US-07-689-008-5
 ; Sequence 5, Application US/07689008
 ; Patent No. 5268274
 ; GENERAL INFORMATION:
 ; APPLICANT: Ben-Bassat, Arle
 ; APPLICANT: Calhoun, Roger D
 ; APPLICANT: Fear, Anna L
 ; APPLICANT: Gelfand, David H
 ; APPLICANT: Meade, James H
 ; APPLICANT: Tal, Rony
 ; APPLICANT: Wong, Hing
 ; APPLICANT: Benitman, Moshe
 ; TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE
 ; TITLE OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McCutchen, Doyle, Brown & Eneersen
 ; STREET: Three Embarcadero Center
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA

```

      ZIP: 94111
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.24
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/07/689,008
      FILING DATE: 19910422
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 337,194
      FILING DATE: 12-APR-1989
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 496,236
      FILING DATE: 23-MAR-1990
      ATTORNEY/AGENT INFORMATION:
      NAME: Murphy, Lisabeth Felix
      REGISTRATION NUMBER: 31547
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (415) 393-2000
      TELEFAX: (415) 393-2286
      TELEX: 340817 MACPAG SFO
      INFORMATION FOR SEQ ID NO: 5:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 3957 base pairs
      TYPE: NUCLEIC ACID
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      US-07-689-008-5

```

US-10-069-544-2 (1-204) x US-07-689-008-5 (1-3957)

```

QY      26 ProProlAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeu 45
      847 CCCCCCGGTGTCGCGCGCGACAGCGCGGCTTGGCGCCACAGAGTTACACAGACCTT 906
      46 AspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArg 65
      907 AACCGCGGCGCTTGGCGCGACCGACGATCTTCCAGTCGCGCTTCAGATCAATTC 966
      66 GlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAspAlaThr 85
      967 CATGATGCTGATTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1023
      86 HisGlnThrLysAlaGlnArgLeuPheGluLysAlaIleGluLeuAsnProLysAspMet 105
      1024 -----GCCGAGCGCGCGCTATTTGAAGAGCGAGCGCGCCAGCCACCAAGCCGCC 1077
      106 GlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeu 125
      1078 GATCGCTG----- 1086
      126 IleTyrPheAspLysProSerArgAla---IleGlyTyrGlnGlyArgValAlaIle 144
      1087 -----CGCCCGCGCTTGGCGGACATGCGCGTCACAGCGGAGATATGCTTCGTT 1134
      145 GluAsnMetAlaTyrIleTyrThrHisGlnTyrGlnAlaAlaLys 159
      1135 CGCAGATTG-----ATTGCGCGCATCATATACAGAGCGCAG 1173

```

RESULT 15

US-07-689-008-1
 ; Sequence 1, Application US/07689008

Patent No. 5268274
GENERAL INFORMATION:
APPLICANT: Ben-Bassat, Arle
APPLICANT: Calhoun, Roger D
APPLICANT: Fear, Anna L
APPLICANT: Gelfand, David H
APPLICANT: Meade, James H
APPLICANT: Tal, Ronny
APPLICANT: Wong, Hing
APPLICANT: Benziman, Moshe
TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE
EXPRESSION OF CELLULOSE SYNTHASE OPERON
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: McCutchen, Doyle, Brown & Eversen
STREET: Three Embarcadero Center
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/689,008
FILING DATE: 19910422
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 337,194
FILING DATE: 12-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 496,236
FILING DATE: 23-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Lisabeth Feix
REGISTRATION NUMBER: 31547
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 393-2000
TELEFAX: (415) 393-2286
TELEX: 340817 MACPAG SFO
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9540 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(328..2589, 2594..4999, 5005..8961, 8964
LOCATION: ...9431)
OTHER INFORMATION:
US-07-689-008-1
Alignment Scores:
Pred. No.: 2.72 Length: 9540
Score: 83.50 Matches: 31
Percent Similarity: 37.78% Conservative: 20
Best Local Similarity: 22.96% Mismatches: 57
Query Match: 8.01% Indels: 27
DB: 1 Gaps: 4
US-10-069-544-2 (1-204) x US-07-689-008-1 (1-9540)
QY 26 PropolysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeu 45
DB 5851 CCCCCGGTGTCCCGCGGACCAAGCCGGGCTTGGCCGCGGACGAGGTACAGCAGCTT 5910
QY 46 AapmetGlyLysLeuAspGlnAlaIleGlnGlnLeuAspAlaIleLeuSerAlaAspArg 65
DB 5911 AACGGGCGCCGCTTGTCCGCGACGCGAGCGAGCTTTCAGTCGGCGTTCAGATCAATTCC 5970

QY 66 GlnPheAlaProAlaTyrArgThrIleuAlaIleValTyrGlnAlaSerGluAspAlaThr 85
DB 5971 CATGATGCTGATTCGCTTGGCGATGGCGCTCGTAAGCAATCGCGAGCGGATACCC--- 6027
QY 86 HisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMet 105
DB 6028 -----GGGAGCGCGCGCTATTGTAAGAACGATGCGCGCCAGCCCAAGACCGCC 6081
QY 106 GlnSerTyrMetAlaTyrGlyPheTyrIleuValGlnMetGlyAspLeuSerGlyAlaLeu 125
DB 6082 GATCGCTGG----- 6090
QY 126 IleTyrPheAspLysProSerArgAla---IleGlyTyrGlnGlyArgValAlaIle 144
DB 6091 -----CCCGCGCGCTTGGCGGATGGCGCGGCGGAGGAGTATGCTTCCTT 6138
QY 145 GluAsnMetAlaTyrIleTyrThrIleGlnTyrGlnAlaValLys 159
DB 6139 CGCCAGTTG-----ATTGCCGCCCATCATATACCGAGGCCAAG 6177
RESULT 16
US-09-134-001C-322
Sequence 322, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: 1998-08-13
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 322
LENGTH: 30549
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-322
Alignment Scores:
Pred. No.: 16.5 Length: 30549
Score: 83.50 Matches: 49
Percent Similarity: 37.72% Conservative: 37
Best Local Similarity: 21.49% Mismatches: 74
Query Match: 8.01% Indels: 68
DB: 4 Gaps: 9
US-10-069-544-2 (1-204) x US-09-134-001C-322 (1-30549)
QY 21 GlnSerThrProIleProPolysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIle 40
DB 11410 GAAGCAAAAGTATTTTAAACAAATGACATCTCAAGTATCAAGACGTCGCCGTTA 11469
QY 41 -----AlaIleSerLeuLeu----- 45
DB 11470 AACAAATAAAGCTGTCAACTGAATTAGATTAACGATGACGATGCTTAAATAA 11529
QY 46 ---AapmetGlyLysLeuAspGlnAlaIleGlnGlnLeuAspAlaIleLeuSer----- 62
DB 11530 GAGAAATATATGATTCGCTTGGTCAAGCAACACCACTTCAACAAATTGTTAATGAAGTA 11589
QY 62 ----- 62
DB 11590 GATCAACACAAAGCATGACACAGATATGCTAATACATAATCAAAAAACGTGA 11649
QY 63 AlaAspArgGlnPheAlaProAlaTyrArgThrIleuAlaIleValTyrGlnAlaSerGlu 82
DB 11650 GCTGAAGATGAATATCAAAAAAGCTCAACAATCTTAACAATGCGATGCGCATGACGCA 11709

QY 83 AspaLatHrHieGlnThrLysAlaGlnArgLeuPheGluLysAlaIleGlnLeuAsnPro 102
 |||:|||||
 DB 11710 CAATATCTACGAAAGAAATAGTAATATCAAGGATTAATAGCATTAACAAAGCCAAA 11769
 QY 103 LysAspMet-----GlnSerTyrMetAspTyrGlyPhe---TyrLeuValGlnMet 118
 |||:|||||
 DB 11770 AACGATTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11829
 QY 119 GlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGlu 138
 |||:|||||
 DB 11830 GTTGATACAAATGGT-----AAAAACCTGCTAGT----- 11859
 QY 139 GlyArgValAlaAlaIleGlnAsnMetAlaTyrIleTyrHisGlnTyrGluAla 158
 |||:|||||
 DB 11860 -----ATTCAACAAATACCAAGCTGCT 11880
 QY 159 LysSerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeu--- 177
 |||:|||||
 DB 11881 CGACAGCTATTGACGCAATACCAATACCAATACCAATACCAATACCAATACCAAT 11940
 QY 178 IleSerGlyThrGlnHisAspGluIle-LysLysSerTyrAspLysLeu---LeuSerAs 196
 |||:|||||
 DB 11941 AATAGTAACCTTCAGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 12000
 QY 196 pTyrLysLeuLeuSerAspTyr 203
 |||:|||||
 DB 12001 CTTAAAGTTAATGACGCAATTC 12022
 RESULT 17
 US-07-618-946B-24
 ; Sequence 24, Application US/07618946B
 ; Patent No. 5453371
 ; GENERAL INFORMATION:
 ; APPLICANT: YUJI SHIBANO et al.
 ; TITLE OF INVENTION: Bacterial Collagenase Gene
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Displaywrite
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/618, 946B
 ; FILING DATE: 19901127
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX: 202-371-8856
 ; TELEX: 440293 WNDP UI
 ; INFORMATION FOR SEQ ID NO: 24:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2217 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE:
 ; HYPOTHETICAL:
 ; ANTI-SENSE:
 ; FRAGMENT TYPE:

; ORIGINAL SOURCE:
 ; ORGANISM:
 ; STRAIN:
 ; INDIVIDUAL ISOLATE:
 ; DEVELOPMENTAL STAGE:
 ; HAPLOTYPE:
 ; TISSUE TYPE:
 ; CELL TYPE:
 ; LINE:
 ; ORGANELLE:
 ; IMMEDIATE SOURCE:
 ; LIBRARY:
 ; CLONE:
 ; POSITION IN GENOME:
 ; CHROMOSOME/SEGMENT:
 ; MAP POSITION:
 ; UNITS:
 ; FEATURE:
 ; NAME/KEY:
 ; LOCATION:
 ; IDENTIFICATION METHOD:
 ; OTHER INFORMATION:
 ; PUBLICATION INFORMATION:
 ; AUTHORS:
 ; TITLE:
 ; JOURNAL:
 ; VOLUME:
 ; ISSUE:
 ; PAGES:
 ; DATE:
 ; DOCUMENT NUMBER:
 ; FILING DATE:
 ; PUBLICATION DATE:
 ; RELEVANT RESIDUES IN SEQ ID NO:
 ; US-07-618-946B-24
 Alignment Scores:
 Pred. No.: 0.451 Length: 2217
 Score: 82.00 Matches: 44
 Percent Similarity: 43.45% Conserves: 29
 Best Local Similarity: 26.19% Mismatches: 57
 Query Match: 7.86% Indels: 38
 DB: 1 Gaps: 11
 US-10-069-544-2 (1-204) x US-07-618-946B-24 (1-2217)
 QY 30 AsnProGlnLeuAlaGlnIle-----ArgThrGlnIleAlaIleSerLeuAsp 46
 |||:|||||
 DB 532 AACGACAAATTTGTGCAATATTTGGCAACCAACGACCTTTGCCAAGCTTTAGGCGAT 591
 QY 47 MetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaLeuSerAlaAspArgGln 66
 |||:|||||
 DB 592 TTTCCT-----CTAAGGGCGCTCATCATGCGTGTGAAGATGAG 630
 QY 67 Phe-----AlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAspAla 84
 |||:|||||
 DB 631 TTATGCGCGCAATGCGGCGGAGAGCTCGGCGCTGACCAAGATATACGGTAAACGG 690
 QY 85 ThrHisGlnThrLysAlaGln-----ArgLeuPheGluLys----- 96
 |||:|||||
 DB 691 AGTTCTGTTGTGAAGGTACGCTGATGATGATGATGATGATGATGATGATGATGATGAT 750
 QY 97 -----AlaIleGlnLeuAsnProLysAspMetGlnSerTyr-----MetAsp 110
 |||:|||||
 DB 751 GGTGACGGGTTTGGCTTTCGCGGGGAGCACCGCTCATATTACGAGATTGTAGTAG 810
 QY 111 TyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle-----Tyr 127
 |||:|||||
 DB 811 TTGGAATTTGTAAATTTGCAACT---GAGCTAAAGGCTGTGCTATCGCAACTTAT 867
 QY 128 PheAspLysPro-----SerArgAlaIleGlyTyrGlnGlyArgValVal 142
 |||:|||||
 DB 868 ACTTAGACCCGACCAATCCGAATTTTGTCTCAGATATATACCAAGACGACACGCGCGCC 927

Qy 143 AlaileglaaenMetalatyrile-----TyrTyrlhsgln---TyrGlualaalaLys 159
Db 928 GCATGTTCTAAATAATGCGTTACGAAGAGGTTACTTCTCATCTGATTGAAGAACTGGTGA 987
Qy 160 SerProthrlysaaspTyraen 167
Db 988 CAGCCAGTAAAGATGACCAACAT 1011

RESULT 18
US-07-618-946B-25
; Sequence 25, Application US/07618946B
; Patent No. 5453371
; GENERAL INFORMATION:
; APPLICANT: YUJI SHIBANO et al.
; TITLE OF INVENTION: Bacterial Collagenase Gene
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618, 946B
; FILING DATE: 19901127
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX: 440293 WNDP UI
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2442 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHEICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; ORGANELLER:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:

LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:

US-07-618-946B-25

Alignment Scores:
Pred. No.: 0.524 Length: 2442
Score: 82.00 Matches: 44
Percent Similarity: 43.45% Conservative: 29
Best Local Similarity: 26.19% Mismatches: 57
Query Match: 7.86% Indels: 38
DB: 1 Gaps: 11

US-10-069-544-2 (1-204) x US-07-618-946B-25 (1-2442)

Qy 30 AsnProGlnLeuAlaGlnIle-----ArgThrGlnIleAlaIleSerLeuLeuAsp 46
Db 757 AACGAGCAATTGTGCAATTAATGGCAACCAACGACCTTGCAAGCTTTAGGCGAT 816
Qy 47 MetGlyLysLeuAspGlnAlaIleGlnLeuAspAlaIleAspArgGln 66
Db 817 TTTGCT-----CTAAGGCGTCATCAATCGTCTGAAGATGAG 855
Qy 67 Phe-----AlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAspAla 84
Db 856 TTTATGCGCGCGAATGCGGCGGAGAGCTCGGCGGTGACCAAGTATACGGTAAACGCG 915
Qy 85 ThrIleGlnThrLysAlaGln-----ArgLeuPheGluLys----- 96
Db 916 AGTTCTGTTGAGAGAGTCACTGATCGAATCTTTGAACGATGAATGATGATGTCGG 975
Qy 97 -----AlaIleGluLeuAspProLysAspMetGlnSerTyr-----MetAsp 110
Db 976 GGTACGCGGCTTTGGCTTTGGCGCGGACACCGCTCATATTACGAGATTGATGATGAG 1035
Qy 111 TyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle-----Tyr 127
Db 1036 TTCGGAATTGTAAATTTCGAAACT---GAGCTAAAGGCTTGATGTCGCAACTTAT 1092
Qy 128 PheAspLysPro-----SerArgAlaIleGlyTyrGluGlyArgValVal 142
Db 1093 ACTGTAGCCCGACATCCGAAATTTGTCTCAGATATGACGAAGCAACACCGCGCC 1152
Qy 143 AlaileglaaenMetalatyrile-----TyrTyrlhsgln---TyrGlualaalaLys 159
Db 1153 GCATGTTCTAAATAATGCGTTACGAAGAGGTTACTTCTCATCTGATTGAAGAACTGGTGA 1212
Qy 160 SerProthrlysaaspTyraen 167
Db 1213 CAGCCAGTAAAGATGACCAACAT 1236

RESULT 19
US-07-618-946B-2
; Sequence 2, Application US/07618946B
; Patent No. 5453371
; GENERAL INFORMATION:
; APPLICANT: YUJI SHIBANO et al.
; TITLE OF INVENTION: Bacterial Collagenase Gene
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack

```

Pred. No.: 1.15 Length: 4054
Score: 82.00 Matches: 44
Percent Similarity: 43.45% Conservative: 29
Best Local Similarity: 26.19% Mismatches: 57
Query Match: 7.86% Indels: 38
DB: 1 Gaps: 11

US-10-069-544-2 (1-204) x US-07-618-946B-2 (1-4054)

QY 30 AsnProGlnLeuAlaGlnIle-----ArgThrGlnIleAlaIleSerLeuLeuAsp 46
Db 2093 AACGAGCAATTGTGGCAATTAATTGGACCAACGACCTTGGCAAGCTTTAGCCGAT 2152
QY 47 MetGlyIysLeuAspGlnAlaIalysGlnIleuAspAlaIaleSerAlaAspArgGln 66
Db 2153 TTTCCT-----CTAAGGCGCTCATCATTCATTCGGTCTGAAGATAG 2197
QY 67 Phe-----AlaProAlaTyrArgThrLeuAlaIalysValTyrGlnAlaSerGluAspAla 84
Db 2192 TTATAGCGCGGAAATGCGGGCGAGAGCTCGGGCGCTGACCAAGATATACGGTAAACCG 2251
QY 85 ThrHisGlnThrIylsAlaGln-----ArgLeuPheGlyIys----- 96
Db 2252 AGTCTCTGTGGAGAGACAGCTAGTGGAACTTTGAAACGATGAATAATGATGCTGG 2311
QY 97 -----AlaIleGluLeuAsnProLysAspMetGlnSerTyr-----MetAsp 110
Db 2312 GGTGACGCGGTTTGCGCTTCGGCGGCGGACACCGCTCATTTACGAGATTGTAGTGA 2377
QY 111 TyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle-----Tyr 127
Db 2372 TTGGAAATTTGTAATTTGGAACCT---GAGCTAAAGGCTTGTCCTATCGCAACTAT 2422
QY 128 PheAspLysPro-----SerArgAlaIleGlyTyrGluGlyArgValVal 142
Db 2429 ACTTGTAGCCCAACATCCGAATTTGTCTCAGAAATATGACCAAGACACACCGCGCC 2488
QY 143 AlaIleGluAsnMetAlaTyrIle-----TyrTyrHisGln---TyrGluAlaIalys 159
Db 2489 GCAATGTTCTTAAATGCGTTTACCAAGAGGTTTACTTTCATCATGCTATTGAACGTGTGA 2544
QY 160 SerProThrLysAspAspTyrAsn 167
Db 2549 CAGCCAGTAAAGATGACCACAACT 2572

RESULT 20
US-09-453-702B-206/C
; Sequence 206, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli 0157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinkney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 43360
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-09-453-702B-206

Alignment Scores:
Pred. No.: 45.6 Length: 43360
Score: 82.00 Matches: 35
Percent Similarity: 50.00% Conservative: 23
Best Local Similarity: 30.17% Mismatches: 43
Query Match: 7.86 Indels: 15
DB: 4 Gaps: 7

US-10-069-544-2 (1-204) x US-09-453-702B-206 (1-43360)

QY 1 Metysileargvalylvtrp---Prometvalmetcalametgyleuvalleuserala 19
Db 24844 TTGAAGCCTTAAAGTGTGCTGAGTATGTGACAGAAATTATATCTATTCGTA 24785
QY 20 Cysglnserthrproileprolyasabn-Proglnleuaglnleargthgl 39
Db 24784 AAGCAGACCAACCGATT-----GAGAGATCAATGCTTAAGATTAACTTCA 24737
QY 39 nllalalileserleuaspmetgilyleuaspqlnalalyglnleuaspal 59
Db 24736 G-----ATAAAGAGGTGAT--GGTAACTGATGACTGCGCAATCCATTAAAGA 24686
QY 59 aalaleuseralaasparglnphealaproalatyargthrlleualalyval----- 77
Db 24685 GTTCATTCCTCTAAGCAGCATATTGTCATCGCTGGAATTGGATTAAGTCTGTT 24626
QY 78 -----Tyrqlnalasergluasparlatrhieglnthrllysalaglnatgyleuphegl 95
Db 24625 GTATTAATATCAGATTAAACAAATGCAATTGATGACAAAG---CAGCGATTGTATGA 24569
QY 95 ulysalalilegluleuasnprolyasapmetglnserlyrmetasp 110
Db 24568 GAAAAAATCTTCACCTTGG---AAGGAAAGCGTTCAATTGCTAAGT 24526

RESULT 21
US-09-453-702B-261/c
Sequence 261, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plunkney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 261:
SEQUENCE CHARACTERISTICS:
LENGTH: 45325 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 261:
US-09-453-702B-261

Alignment Scores:
Pred. No.: 48.9 Length: 45325
Score: 82.00 Matches: 35
Percent Similarity: 50.00% Conservative: 23
Best Local Similarity: 30.17% Mismatches: 43
Query Match: 7.86 Indels: 15
DB: 4 Gaps: 7

US-10-069-544-2 (1-204) x US-09-453-702B-261 (1-45325)

QY 1 Metysileargvalylvtrp---Prometvalmetcalametgyleuvalleuserala 19
Db 25746 TTGAAGCCTTAAAGTGTGCTGAGTATGTGACAGAAATTATATCTATTCGTA 25687
QY 20 Cysglnserthrproileprolyasabn-Proglnleuaglnleargthgl 39
Db 25686 AAGCAGACCAACCGATT-----GAGAGATCAATGCTTAAGATTAACTTCA 25639
QY 39 nllalalileserleuaspmetgilyleuaspqlnalalyglnleuaspal 59
Db 25638 G-----ATAAAGAGGTGAT--GGTAACTGATGACTGCGCAATCCATTAAAGA 25588
QY 59 aalaleuseralaasparglnphealaproalatyargthrlleualalyval----- 77
Db 25587 GTTCATTCCTCTAAGCAGCATATTGTCATCGCTGGAATTGGATTAAGTCTGTT 25528
QY 78 -----Tyrqlnalasergluasparlatrhieglnthrllysalaglnatgyleuphegl 95
Db 25527 GTATTAATATCAGATTAAACAAATGCAATTGATGACAAAG---CAGCGATTGTATGA 25471
QY 95 ulysalalilegluleuasnprolyasapmetglnserlyrmetasp 110
Db 25470 GAAAAAATCTTCACCTTGG---AAGGAAAGCGTTCAATTGCTAAGT 25428

RESULT 22
US-09-328-111-842/c
Sequence 842, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.

APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Derti, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 842
LENGTH: 609
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(609)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-842
Alignment Scores:
Pred. No.: 0.0972 Length: 609
Score: 80.50 Matches: 21
Percent Similarity: 50.72% Conservative: 14
Best Local Similarity: 30.43% Mismatches: 25
Query Match: 7.72% Indels: 9
DB: Gaps: 2
US-10-069-544-2 (1-204) x US-09-328-111-842 (1-609)
QY 47 MetGlyLeuAspGlnAlaLeuAspAlaLeuSerAlaAspArgGln 66
DB 410 TTGGGCGCTGGAGAGCGCAAGCATGTTTGAAGCATTTAGACGCAAGC 351
QY 67 PheAlaProAlaTyrArgThrLeuAlaValTyrGlnAlaSerGluAsp 83
DB 350 TTTCGAGTAGCTTGGAGTATCTTGGCTGTTTCAATGACCAAGGAAATTGGCTT 291
QY 84 AlaThrHisGlnThrLysAlaGlnArgLeuPheGlnLysAlaLeuSerGluAsp 103
DB 290 GCAATTCATCAC-----TTTGAAGGCTGTCACTTGACCCCAAC 249
QY 104 AspMetGlnSerTyrMetLysPheTyrGly 112
DB 248 TTTCGATGCTTAATCAATTAGGA 222
RESULT 23
US-09-051-019-1
Sequence 1, Application US/09051019
Patent No. 6103229
GENERAL INFORMATION:
APPLICANT: KAHMANN, Regine and QUADBECK-SEBGER, Claudia
TITLE OF INVENTION: Regulatory gene from Ustilago maydis
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Keil & Weikauf
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
COMPUTER: IBM AT-compatible, Pentium processor
OPERATING SYSTEM: Windows 98
SOFTWARE: Wordperfect version 6.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/051,019
FILING DATE: 31-MAR-1998
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8931 base pairs
TYPE: Nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-051-019-1
Alignment Scores:
Pred. No.: 6.28 Length: 8931
Score: 80.50 Matches: 39
Percent Similarity: 39.05% Conservative: 27
Best Local Similarity: 23.08% Mismatches: 48
Query Match: 7.72% Indels: 55
DB: Gaps: 7
US-10-069-544-2 (1-204) x US-09-051-019-1 (1-8931)
QY 54 LysGlnGlnLeuAspAlaAlaLeu-----SerAlaAspArgGlnPheAla 68
DB 5141 AGGCAAGCTTGGAGCGCTTCGTCAGCATGCCAATCCACTTCGATCGAAGTCCGAA 5200
QY 69 ProAlaTyrArgThrLeuAlaLysVal-----TyrGlnAlaSerGluAspAla 84
DB 5201 GCTTGTATGGCTTGATGGAGAGCTCAGACGCTTCACTTGACGCGCTGAGATTGCA 5260
QY 85 ThrHisGlnThrLysAlaGlnArgLeu-----Phe 94
DB 5261 TCGCTTGCACATTAAGCGCAAGAGCTCGAGATTGATGCGCGGTGACGAGTCTTA 5320
QY 95 GluLysAlaLeuGluLeuAspProLysAspMetGlnSerTyrMetAspTyrGlyPheTyr 114
DB 5321 CAACAGGCTGACGAGACTAATCTCAAGACTGTGAAGACATCTGACGCTGCGACGCTT 5380
QY 115 Leu-----ValGlnMetGlyAspLeuSerGlyAlaLeuLeuTyrPheAspLys 130
DB 5381 CTCATGTGACGCGCTCAATCAAGAGCTTCC----- 5416
QY 131 ProSerArgAlaLeuGlyTyrGlnLysValAlaAlaLeuGluAspMet----- 147
DB 5417 -----GACTATGTGCGAGCTGCAAGTGCAGAGTCAAGAGTCAAGATCG 5461
QY 148 -----AlaTyrLleTyrThrHisGlnTyrGlnAlaAlaLysSerProThrLysAsp 165
DB 5462 TTCGACATATCTCTATTACCAAGAA----- 5488
QY 166 TyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeuLysSerGlyThrGlnHisAspGlu 185
DB 5489 -----GTTGCGGAAGCTGTGATGCGCCCGACACTGTGTCTCAAGATCAGAG 5539
QY 186 IleLysLysSerTyrAspLysLeu 194
DB 5540 CTGCGCAAGAACTTGACGACAGACTC 5566
RESULT 24
US-09-147-236-1
Sequence 1, Application US/09147236A
Patent No. 6316251
GENERAL INFORMATION:
APPLICANT: TONOUCHI, Naoto
APPLICANT: TSUCHIDA, Takayasu
APPLICANT: YOSHINAGA, Fumihito
APPLICANT: TAHARA, Naoki
APPLICANT: HAYASHI, Takahisa
TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE
FILE REFERENCE: 6537-011-OPCT
CURRENT APPLICATION NUMBER: US/09/147,236A
CURRENT FILING DATE: 1999-04-08
EARLIER APPLICATION NUMBER: PCT/JP97/03633
EARLIER FILING DATE: 1997-10-09

```

; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 16836
; TYPE: DNA
; ORGANISM: Acetobacter xylinum
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (869)..(1891)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3101)..(5368)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (5373)..(7778)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7784)..(11761)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (11764)..(12231)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (12448)..(14652)
;   FEATURE:
;   OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, t, or
;   OTHER INFORMATION: c
US-09-147-236-1
```

```

Alignment Scores:
Pred. No.: 16.8      Length: 16836
Score: 80.50      Matches: 31
Percent Similarity: 32.43%      Conservative: 17
Best Local Similarity: 20.95%      Mismatches: 47
Query Match: 7.72%      Indels: 53
DB: 4      Gaps: 4

US-10-069-544-2 (1-204) x US-09-147-236-1 (1-16836)
```

```

QY 26 ProProlysaAaAaAaProGlnLeuAlaGlnIleAArgThrGlnIleAlaIleSerLeuLeu 45
DB 8630 CCCCCCGGCGCCCGCGCAAAAGCGGGCTTCGCGCGCAGCCGCTTACCGACGAGCTC 8689
QY 46 AspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArg 65
DB 8690 AACGGGGCGCTTCCTCCGCTCCGACAGTCC----- 8722
QY 66 GlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAspAlaThr 85
DB 8722 ----- 8722
QY 86 HisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMet 105
DB 8723 -----TTCCAGTGGCGGCTTCAGATCACTCCCATGATGCT 8758
QY 106 GlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeu 125
DB 8759 GATTCACTGGGTGGCATGGGCTTCGTGACATGGCGAGGGGATACCGCGAGCGGAC 8818
QY 126 IleTyrPheAsp-----LysProSer 132
DB 8819 CCGTATTTTGAAGAGCGCATGGCGCGACCCCAAGACTGCCGATCGCTGGCGCGCGG 8878
QY 133 ArgAla---IleGlyTyrGlnGlyArgValAlaAlaIleGlnAsnMetAlaTyrIleTyr 151
DB 8879 CTTCGGGCGATGGCGGTACGCGGGGACTATATGCCGCTTCGCCAGTTG-----ATTGCC 8932
QY 152 TyrHisGlnTyrGlnAlaAlaLys 159
DB 8933 GCCCATCATATACCGAGCCAG 8956

RESULT 25
US-09-147-236-10
```

```

; Sequence 10, Application US/09147236A
; Patent No. 6316251
; GENERAL INFORMATION:
;   APPLICANT: TONOUCHI, Naoto
;   APPLICANT: TSUCHIDA, Takayasu
;   APPLICANT: YOSHIMAGA, Fumihito
;   APPLICANT: TAHARA, Naoki
;   APPLICANT: HAYASHI, Takahisa
;   TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE
;   FILE REFERENCE: 6537-011-0PCT
;   CURRENT APPLICATION NUMBER: US/09/147, 236A
;   EARLIER FILING DATE: 1999-04-08
;   EARLIER APPLICATION NUMBER: PCT/JP97/03633
;   NUMBER OF SEQ ID NOS: 12
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 10
;   LENGTH: 16836
;   TYPE: DNA
;   ORGANISM: Acetobacter xylinum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1891)..(2922)
;   FEATURE:
;   OTHER INFORMATION: Nucleotide sequence is the same as SEQ ID NO:1
;   OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, c, or
;   OTHER INFORMATION: t
US-09-147-236-10
```

```

Alignment Scores:
Pred. No.: 16.8      Length: 16836
Score: 80.50      Matches: 31
Percent Similarity: 32.43%      Conservative: 17
Best Local Similarity: 20.95%      Mismatches: 47
Query Match: 7.72%      Indels: 53
DB: 4      Gaps: 4

US-10-069-544-2 (1-204) x US-09-147-236-10 (1-16836)
```

```

QY 26 ProProlysaAaAaAaProGlnLeuAlaGlnIleAArgThrGlnIleAlaIleSerLeuLeu 45
DB 8630 CCCCCCGGCGCCCGCGCAAAAGCGGGCTTCGCGCGCAGCCGCTTACCGACGAGCTC 8689
QY 46 AspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArg 65
DB 8690 AACGGGGCGCTTCCTCCGCTCCGACAGTCC----- 8722
QY 66 GlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAspAlaThr 85
DB 8722 ----- 8722
QY 86 HisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMet 105
DB 8723 -----TTCCAGTGGCGGCTTCAGATCACTCCCATGATGCT 8758
QY 106 GlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeu 125
DB 8759 GATTCACTGGGTGGCATGGGCTTCGTGACATGGCGAGGGGATACCGCGAGCGGAC 8818
QY 126 IleTyrPheAsp-----LysProSer 132
DB 8819 CCGTATTTTGAAGAGCGCATGGCGCGACCCCAAGACTGCCGATCGCTGGCGCGCGG 8878
QY 133 ArgAla---IleGlyTyrGlnGlyArgValAlaAlaIleGlnAsnMetAlaTyrIleTyr 151
DB 8879 CTTCGGGCGATGGCGGTACGCGGGGACTATATGCCGCTTCGCCAGTTG-----ATTGCC 8932
QY 152 TyrHisGlnTyrGlnAlaAlaLys 159
DB 8933 GCCCATCATATACCGAGCCAG 8956

RESULT 26
```

US-09-221-017B-1028
Sequence 1028, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FORRESTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P11182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA: P11546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA: P22911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ. ID NO: 1028:
SEQUENCE CHARACTERISTICS:
LENGTH: 821 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1...821
US-09-221-017B-1028
Alignment Scores:
Pred. No.: 0.211 Length: 821
Score: 79.50 Matches: 34
Percent Similarity: 40.28% Conservative: 24
Best Local Similarity: 23.61% Mismatches: 73
Query Match: 7.62% Indels: 13
Gaps: 4
US-10-069-544-2 (1-204) x US-09-221-017B-1028 (1-821)
QY 51 AspglAlalysglInleuaspAlaAlaleuSerAlasprgInlphAlaProAla 70
DB 148 GACCAAGCCGACGATTATTAACAAGAAAGCATGTAGAGGACAGATCCGATCATGCG----- 201

QY 71 TyrArgThrleuAlalysValTyrGlnAlaSerGluaspAlaThrHisGln---ThrLys 89
DB 202 ---AATACTCTTGGGAATTATGCCAATTCTTATGTATATATGCTACAGATTAGACCA 258
QY 90 AlaGlnArgleuPhegluYleAlaIleGluLeuasnProLysaspMetGlnSerTyrMet 109
DB 259 GCCGAGGGGTATTACAGAGAGACATTAGAGGACATCCGAAGAAAGCAATTATCTTGGC 318
QY 110 AspTyrGlyPheTyrleuValGlnMetGlyLysaspSerGlyAlaLeuIleTyrPheasp 129
DB 319 AATTATGCTTATTCTTA-----AATATATTCGTCCGCTTAC----- 357
QY 130 LysProSerArgAlaIleGlyTyrGluGlyArgValAlaIleGluAsnMetAlaTyr 149
DB 358 -----GACCAAGCCGAGCGCTATTACAGGACGATTAGAGTAGATCCAGAGATGCC 411
QY 150 IleTyrThrIsglnTyrGluAlaAlaLysSerProThrLysaspAspTyrAsnAsnAla 169
DB 412 AATAAGCTTGGCAATTATGCTCATTTCTCATTCAGTGTCCGGGGGATTCMAACGTGCT 471
QY 170 LysSerAlaLeuGluArgAlaLeuIleSerGlyThrGlnHisaspGluIleLysLysSer 189
DB 472 GATAGCTTGATTCAGCAAGCATTCGAGATGCCGATTAATGACGAGACGAAACCTTG 531
QY 190 TyrAspLysLeu 193
DB 532 CAGGCCAAGTTR 543
RESULT 27
US-07-792-865D-2
Sequence 2, Application US/07792865D
Patent No. 5646247
GENERAL INFORMATION:
APPLICANT: John W. Barnwell, Mary W. Galinski,
APPLICANT: Samuel P. Wertheimer
TITLE OF INVENTION: MEROZOITE ANTIGENS LOCALIZED AT
TITLE OF INVENTION: THE APICAL END OF THE PARASITE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby P.C.
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
COMPUTER: IBM or IBM-compatible
OPERATING SYSTEM: PC/MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/792,865D
FILING DATE: 19911004
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US90/01849
FILING DATE: April 3, 1990
ATTORNEY/AGENT INFORMATION:
NAME: Jackson, Arthur
REGISTRATION NUMBER: 34,354
REFERENCE/DOCKET NUMBER: 5986/14692-US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1985 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
DESCRIPTION:
HYPOTHEICAL: NO

ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: P.vivax
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY: Lambda gt 11 native P.vivax
LIBRARY: DNA expression library
CLONE: 7.2
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: This sequence corresponds to
OTHER INFORMATION: Figure 1B (sheets 1 and 2) in
OTHER INFORMATION: the application, as filed.
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-07-792-865D-2

Alignment Scores:
Pred. No.: 0.833 Length: 1985
Score: 79.50 Matches: 42
Percent Similarity: 38.34% Conservative: 32
Best Local Similarity: 21.76% Mismatches: 58
Query Match: 7.62% Indels: 61
DB: 1 Gaps: 8

US-10-069-544-2 (1-204) x US-07-792-865D-2 (1-1985)

Qy 21 GlnSerThrProIleProProIleProProIleProGlnIleAlaGlnIleArgThrcGlnIle 40
Db 955 GAAAAAGCCAAAGTGAAGAAAAAGAAATCTCGCAACTTAACGATGTTCCACGAG--- 1011
Qy 41 AlaIleSerLeuLeuAspMetClyLysLeuAspGlnAlaIleGlnIleAspAlaIle 60
Db 1012 -----TCGTTATTA-----CAATAGATTAATTCGACGACACAGCTTGACGCGTT 1056
Qy 61 LeuSer-----AlaAspArgGlnPheAlaProAlaTyrrg 72
Db 1057 TTGTCAAACATTGGAAGGCTGAACAACAAATGCACTCAATATTTCGATTCGCGCTGATAAA 1116
Qy 73 ThrLeuAlaIleValTyrcIleAlaSerGluAspAlaIleThr----- 88
Db 1117 TCGATGAAGTCGTTTTCCTTAATGACGATTTGGCTCCGAAAAAATCCCTAGACAAAGTA 1176
Qy 89 ---LysAlaGlnArgLeuPheGlnIleValAlaIleGlnIleAspProLysAspMetGlnSer 107
Db 1177 AAAGCGGCTAAGGAAGATTATGAGAAAATTTGGAACCGCTTCAAAATGAATGAGT--- 1233
Qy 108 TyrMetAspTyrcIlePheTyrcIleValGlnMetClyAspLeuSerGlyAlaLeuIleTyrc 127

Db 1234 -----CGTATTAACTGAGAGAGAGAGCTGACCGAC 1266
Qy 128 PheAspLysProSerArgAlaIleGlyTyrcIleGluIleValValAlaIleGluAspMet 147
Db 1267 ATGACACAA-----AAATTAAGTGAATGAGAAAT--- 1296
Qy 148 AlaTyrcIleTyrcIleGlnIleTyrcIleAlaIleLysSerProThrLysAspTyrrg 167
Db 1297 -----GACTGCTA 1305
Qy 168 AsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGlyThrGlnIleAspGlnIleLys 187
Db 1306 AAATGAGAAACATATATGAGAGAGCTTACTA-----CAAAAGATTAAA 1350
Qy 188 LysSerTyrrgAspLysLeuLeuSerAspTyrrgLeuLeu 200
Db 1351 GAAATCGGATTAAGAGAGAGAGATTATTTGAAATTAGTA 1389

RESULT 28
US-09-221-017B-1087
Sequence 1087, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1087:
SEQUENCE CHARACTERISTICS:
LENGTH: 830 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE: PORPHYROMONAS GINGIVALIS

APPLICANT: Bjvick, Lars
APPLICANT: Sjvbring, Ulf
TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,475
FILING DATE: 11-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: 100084.40201
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1332 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1329
US-08-795-475-5

Alignment Scores:
Pred. No.: 0.613 Length: 1332
Score: 78.50 Matches: 52
Percent Similarity: 39.06% Conservative: 23
Best Local Similarity: 27.08% Mismatches: 66
Query Match: 7.53% Indels: 51
DB: 2 Gaps: 11

US-10-069-544-2 (1-204) x US-08-795-475-5 (1-1332)
Qy 29 AenAenProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGly 48
Db 46 AACAAATCCCGCAATCAAAATATACGTTTACGT-----CAC 81
Qy 49 LysLeuAspGlnAlaLysGlnLeuLeuAspAlaAlaLeuSer---AlaAspArgGlnPhe 67
Db 82 GAAACCAAGACTTAAGACGAGATTAGAGATGCAATGCAAGCTTGCAAGAAAGATT 141
Qy 68 AlaPcoAlaIstArgThrLeuAlaLysValIstArgGlnAlaSerGlnAspAlaThrHisGln 87
Db 142 AAGAGAGCT---GAAGAACTTGAAAAAGCAAAAGCCTTGAAGAC----- 186
Qy 88 ThrLysAlaGlnArgLeuPheGlnLysValAlaIleGlnLeuAspProLysAspMetGlnSer 107
Db 187 -----CACGCT-----AAAGATTGAAACTTAATTAAGAACTTACAA--- 225
Qy 108 TyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyr 127
Db 226 ---CAAGACTAT-----GACTTGCAAGAAATCAACAGT 258
Qy 128 PheAspLysProSerArgAlaIleGlyTyrGlnGlyArgValAlaIleGlu----- 145
Db 259 TGGGATGACAAAGACTTGAAAAAGAGATTGAAAGAAAAAGAAAGAGCTCTTGAAATTACGC 318

Qy 146 ---AenMetAlaTyrIleTyrThrGln----- 154
Db 319 ATAGACCAAGCAATCGGACTTACCATAGAGCTACCGCTTAGAAAAAGAGTTAGAGAG 378
Qy 155 -----TyrGlnAlaAlaLysSerProThrLysAspAspTyrAenAsnAlaLys 170
Db 379 AAAAAGAAAGCTCTTGATTAATGATAGACCAAGCAGTCAAGACTATTAATAGAGCT--- 435
Qy 171 SerAlaLeuGlnArgAlaLeuIleSerGlyThrGlnHisAspGlnIleLysSerTyr 190
Db 436 AACGCTTAGAAAAAGCTTGAAGAAACGATTACTAGAGAAACAGAGATTAAATGTAAT--- 492
Qy 191 AspLysLeuLeuSerAspTyrLysLeuLeuSerAsp 202
Db 493 -----CTTTAGGCATGCAAAAGCTTGAAGCTTAT 522

RESULT 31
US-08-569-150A-2
Sequence 2, Application US/08569150A
Patent No. 5925804
GENERAL INFORMATION:
APPLICANT: Hoekema, Andreas
APPLICANT: Pen, Jan
APPLICANT: Does, Mirjam P
APPLICANT: Van Den Elzen, Petrus J. M
TITLE OF INVENTION: PRODUCTION OF TREHALOSE IN PLANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ladac & Parry
STREET: 26 West 61st Street
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10023
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/4" disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Wordperfect for Windows
SOFTWARE: Wordperfect 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,150A
FILING DATE: 21-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP/94/02167
FILING DATE: 30-JUNE-1994
ATTORNEY/AGENT INFORMATION:
NAME: Maas, Clifford J.
REGISTRATION NUMBER: 30,086
REFERENCE/DOCKET NUMBER: U-010552-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 708-1890
TELEFAX: (212) 246-8959
TELEX: No. 5925804e
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1446 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
IMMEDIATE SOURCE:
CLONE: 7F11
POSITION IN GENOME:
MAP POSITION: 41-42'
FEATURE:
NAME/KEY: CDS
LOCATION: 19..1446
OTHER INFORMATION: /product= "trehalose phosphate synthase"
OTHER INFORMATION: /gene= "otcA"

US-08-569-150A-2

Alignment Scores:

Pred. No.: 0.697 Length: 1446
 Score: 78.50 Matches: 47
 Percent Similarity: 38.16% Conservative: 32
 Best Local Similarity: 22.71% Mismatches: 65
 Query Match: 7.53% Indels: 63
 DB: 2 Gaps: 12

US-10-069-544-2 (1-204) x US-08-569-150A-2 (1-1446)

```

QY 24 ProileProPolysAsnProGlnLeuAlaGlnIleArgThrGlnIle-----Ala 41
DB 742 CCACTGCCCGCCAAA-----CTGGCGCAACTTAAGCGGAACTGAAAAACGTA 789
QY 42 lIeSerLeuLeuAspMetGlyLysLeuAspGlnAlaIleGlnIleuAspAlaIleu 61
DB 790 CAAATATCTTTCTGTGCAACGGCTGATTTATTCAGAGGTTTCCAGAGCGTTTCTC 849
QY 62 SerAlaAsp-----ArgGln 66
DB 850 GCCTATGAGCGTTCGTGAAAAATATCCGACGATCATGTAAATTCGTTATACCCAG 909
QY 67 PheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnuAspAlaThrHis 86
DB 910 ATTGCACCAACGTGGGTGTGATGTCGCAAGCCTATCAG-----GATATTCGTCAT 960
QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAspProLysAspMetGln 106
DB 961 CAG-----CTCGAAATGAACTGACGAAATTAATGTGTA----- 996
QY 107 SerTyrMetAspTyrGly-----PheTyrLeuValGlnMetGly 119
DB 997 -----TACGGCAATTAGCGTGCAGCCGCTTATTTATTTGATCAG----- 1038
QY 120 AspleuSerGlyAlaLeuIleTyrPheAspLysPro-----SerArgAlaIleGlyTyr 137
DB 1039 -----CATTTGACCGTAAATTACTGTGTAATAATATTCGCTAC 1077
QY 138 Glu-----GlyArgValAlaAlaIleGlnuAspMetAlaTyrIleTyrHisGlnTyr 155
DB 1078 TCTGACGTGGCTTAGTACGACGACGCGGATGACGAGTGAAGTGAAGAAAGATAT 1137
QY 156 GluAlaAlaLysSerProThrLys-----AspAspTyrAsnAsnAla 169
DB 1138 GTTGCTGCTCAGACCGCAATCCGCGGCTTCTGTTCTTTCGCAATTTGCGGAGCG 1197
QY 170 LysSerAlaLeuGlnArgAlaLeuIle-----SerGlyThrGlnHisAspGlnIleLys 188
DB 1198 GCAAAACAGTTTAACGTGGCGTTTATTTTAACTTCAACCCCTACAGATGTCAGAACTTGCAGCT 1257
QY 189 SerTyrAspLysLeuLeuSer 195
DB 1258 GCCGTGATCGTGCATTTGACT 1278

```

RESULT 32
 US-08-274-121B-1
 Sequence 1, Application US/08274121B
 Patent No. 6133034
 GENERAL INFORMATION:
 APPLICANT: Arne Reider Strom
 APPLICANT: Inga Kaasen
 APPLICANT: Olaf Bay Skjervold
 APPLICANT: John McDougall
 TITLE OF INVENTION: Methods and Compositions
 TITLE OF INVENTION: Related to The Production
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Calgene, Inc.
 STREET: 1920 Fifth Street
 CITY: Davis

STATE: CA
 COUNTRY: USA
 ZIP: 95616
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
 COMPUTER: Apple Macintosh
 OPERATING SYSTEM: Macintosh 7.1
 SOFTWARE: Microsoft Word 5.1(a)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/274,121B
 FILING DATE: 12-JULY-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/893,099
 FILING DATE: 27-MAY-1992
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Elizabeth Lassen
 REGISTRATION NUMBER: 31,845
 NAME: Donna E. Scherer
 REGISTRATION NUMBER: 34,719
 REFERENCE/DOCKET NUMBER: CGNE 86(1)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (916) 753-6313
 TELEFAX: (916) 753-1510
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2868 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA

US-08-274-121B-1

Alignment Scores:

Pred. No.: 2.02 Length: 2868
 Score: 78.50 Matches: 47
 Percent Similarity: 38.16% Conservative: 32
 Best Local Similarity: 22.71% Mismatches: 65
 Query Match: 7.53% Indels: 63
 DB: 3 Gaps: 12

US-10-069-544-2 (1-204) x US-08-274-121B-1 (1-2868)

```

QY 24 ProileProPolysAsnProGlnLeuAlaGlnIleArgThrGlnIle-----Ala 41
DB 2167 CCACTGCCCGCCAAA-----CTGGCGCAACTTAAGCGGAACTGAAAAACGTA 2214
QY 42 lIeSerLeuLeuAspMetGlyLysLeuAspGlnAlaIleGlnIleuAspAlaIleu 61
DB 2215 CAAATATCTTTCTGTGCAACGGCTGATTTATTCAGAGGTTTCCAGAGCGTTTCTC 2274
QY 62 SerAlaAsp-----ArgGln 66
DB 2275 GCCTATGAGCGTTCGTGAAAAATATCCGACGATCATGTAAATTCGTTATACCCAG 2334
QY 67 PheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnuAspAlaThrHis 86
DB 2335 ATTGCACCAACGTGGGTGTGATGTCGCAAGCCTATCAG-----GATATTCGTCAT 2385
QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAspProLysAspMetGln 106
DB 2386 CAG-----CTCGAAATGAACTGACGAAATTAAGTGA----- 2421
QY 107 SerTyrMetAspTyrGly-----PheTyrLeuValGlnMetGly 119
DB 2422 -----TACGGCAATTAAGCTGAGCGCGCTTATTTATTTGATCAG----- 2463
QY 120 AspleuSerGlyAlaLeuIleTyrPheAspLysPro-----SerArgAlaIleGlyTyr 137
DB 2464 -----CATTTGACCGTAAATTACTGATGATAAATATTCGCTAC 2502
QY 138 Glu-----GlyArgValAlaAlaIleGlnuAspMetAlaTyrIleTyrThrHisGlnTyr 155

```

Db 2503 TCTGACGTGGCTTAGTACCGCCACTGGCGTAGCGGATGACCTGTAGCAAAAGATAT 2562
Qy 156 GUAUAAUySerProThrllys-----AAspTyRAsnAsnAla 169
Db 2563 GTTGCTGCTCAGACCCAGCCCACTCCGGCGCTTCTTCTTTCGCAATTTCGGGAGCG 2622
Qy 170 LysSerAlaLeuGluArgAlaLeuIle--SerGlyThrGlnHisAspGluIleLyslys 188
Db 2623 GCAACGAGTTTAACCTCGCGCTTATTTGTAACCCCTAGATCGGACGAACTTCAGCT 2682
Qy 189 SerTyRAspLysLeuLeuSer 195
Db 2683 GCGCTGATCGTGCATGTACT 2703
RESULT 33
US-08-928-692-16
Sequence 16, Application US/08928692
GENERAL INFORMATION:
APPLICANT: Brody, Howard
APPLICANT: Yaver, Deborah S.
APPLICANT: Lamas, Michael
APPLICANT: Hansen, Kim
TITLE OF INVENTION: Methods for Modifying the Production of
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESS: No. 59587270 No. 5958727disk of No. 5958727th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,692
FILING DATE: 12-SEPT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4944,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 4700 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-928-692-16
Alignment Scores:
Pred. No.: 5.08 Length: 4700
Score: 78.00 Matches: 29
Percent Similarity: 41.94% Conservative: 23
Best Local Similarity: 23.39% Mismatches: 35
Query Match: 7.48% Indels: 37
DB: 2 Gaps: 5
US-10-069-544-2 (1-204) X US-08-928-692-16 (1-4700)
Qy 7 TrrpProMetValMetAlaMetGlyLeuValLeuSerAla-CysGlnInsertThrProIlePr 26
Db 3604 TGGCGG---ATTACGCTCTGGGATCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCC 3660

Qy 26 OProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAs 46
Db 3661 A-----CCAGAGGAGCAGAGTCGACGAAAGGTCAATTCA-----GA 3696
Qy 46 pMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgG1 66
Db 3697 TATTGGCGTACTGCTCCCGGAGAGACCGAATGTAGCTTCTCTCGCAAGTCCGCGGCT 3756
Qy 66 nPhe----- 67
Db 3757 TACAGAGATCAAGCTCATCACCCGAAAGTAGGCAATCCATCATCGGAGCCATCCTGTGG 3816
Qy 68 ----AlaProAlaTyRArgThrLeu-----AlaLysValTyRGl 79
Db 3817 ACCCTCGCCGCTTTTATGACAGTGGAGCTCGGCGCAGAGGCCATACAACAGATCCTGGC 3876
Qy 79 nAlaSerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGluLysAlaIleG1 99
Db 3877 GACTTCGGAAGATGAACTCACACTGATGTGTATCGGGGTACGTGTGAGGACTTTGA 3936
Qy 99 uLeuAsnPro 102
Db 3937 CTTCAGACT 3946
RESULT 34
US-09-339-972-16
Sequence 16, Application US/09339972
Patent No. 6323002
GENERAL INFORMATION:
APPLICANT: Brody, Howard
APPLICANT: Yaver, Deborah S.
APPLICANT: Lamas, Michael
APPLICANT: Hansen, Kim
TITLE OF INVENTION: Methods for Modifying the Production of
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESS: No. 63230020 No. 6323002disk of No. 6323002th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/339,972
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/928,692
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4944,200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 4700 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-339-972-16
Alignment Scores:
Pred. No.: 5.08 Length: 4700

Score: 78.00 Matches: 29
Percent Similarity: 41.94% Conservative: 23
Best Local Similarity: 23.39% Mismatches: 35
Query Match: 7.48% Indels: 37
DB: 4 Gaps: 5
US-10-069-544-2 (1-204) x US-09-339-972-16 (1-4700)
QY 7 TRPProMetValMetAlaMetGlyLeuValLeuSerAla-CysGlnSerThrProIlePr 26
DB 3604 TGGCCG---ATTCAGCCTCTGGGTATCCCTTATGTTCTTCAGAGAGCCGCTCCC 3660
QY 26 CProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAs 46
DB 3661 A-----CCAGAGCGAGCGAGTCGACGAAAGGTCATTTCA-----GA 3696
QY 46 pMetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaAlaLeuSerAlaAspArgG 66
DB 3697 TATTGGCGTACTGCTCTCCCGGAGAGACCGAATGTTAGTTCTCTGCAAGTCCCGCGCT 3756
QY 66 nPhe----- 67
DB 3757 TACGAGGATCAAGCTCATCACCAGAGTAGCAATCATCATCGGAGCCATCTGTTGG 3816
QY 68 ----AlaProAlaIlyArgThrLeu-----AlaLysValIlyG 79
DB 3817 ACCCTCGCCCGTTTAAATGACAGTGAAGCTCGGCGAAGGCCATACAAACAGATCCTGCG 3876
QY 79 nAlaSerGluAspAlaThrHisGlnThrLysAlaGlnLysLeuPheGluLysAlaIleG 99
DB 3877 GACTTCGGAAGATGAACTCACAGTATGCTGTATGCGGGGTACGTTGAGGACTTTCA 3936
QY 99 uLeuAsnPro 102
DB 3937 CTTCGACGCT 3946
RESULT 35
US-09-047-288-1
Sequence 1, Application US/09047288
GENERAL INFORMATION:
APPLICANT: KUMAR, Sanjay
TITLE OF INVENTION: DRUG BINDING PROTEIN
FILE REFERENCE: GP-50000-D1
CURRENT APPLICATION NUMBER: US/09/047,288
PRIOR FILING DATE: 1998-03-24
PRIOR APPLICATION NUMBER: 08/802,191
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1310
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-047-288-1
Alignment Scores:
Pred. No.: 0.818 Length: 1310
Score: 77.50 Matches: 39
Percent Similarity: 44.30% Conservative: 27
Best Local Similarity: 26.17% Mismatches: 50
Query Match: 7.43% Indels: 33
DB: 4 Gaps: 9
US-10-069-544-2 (1-204) x US-09-047-288-1 (1-1310)
QY 50 LeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
DB 691 ATTGACCGAGCTGAGCGCATCATGAGAGTGCTG-----GGCACACCC 732
QY 70 AlaIyArgThrLeuAlaLysValIlyGlnAlaSerGluAspAlaThrHisGlnThrLys 89
DB 733 AGCCCTGAGGTTCCTGGCAAAATC-----TCTCGGAACACGCGC----- 771

QY 90 AlaGlnArgLeuPheGluLysAlaIleGluLeuAsnPro---LysAspMetGlnSerTyr 108
DB 772 -----GGACATATATCATCAGTCCCTGCCCCCATGCCCCCAAGAGACCTGAGCAGCATC 825
QY 109 MetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPhe 128
DB 826 TTCGTTGAGGCCAACCCCTGGCCAT-----GACTTCCTTGGAAAGATGCTGCTGCTG 879
QY 129 AspLysProSerArgAlaIleGlyTyrGluGlyArgValAlaIleGluAsnMetAla 148
DB 880 GACAGT-----GACCAGAGGCTCATGCTGACGTCGAGGACGCTGCGC 918
QY 149 TyrIleTyrThrHisGlnTyr---GluAlaAlaLysSerProThrLysAspAspTyrAsn 167
DB 919 CAGCCTTACTTCACCCAGTACACAGCCGAGAGTAGCAGCAGGCGCCAGCCATATGAT 978
QY 168 AsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGlyThrGlnHisAspGluIleLys 187
DB 979 GAGAGCGTTGAGGCCAGAGGAGCGCAGCCTG-----GAGGAGTGGAAG 1020
QY 188 Lys---SerTyrAspLysLeuLeuSer 195
DB 1021 GAGCTCACTTACAGGAAGTCTTAGC 1047
RESULT 36
US-08-802-191-1
Sequence 1, Application US/08802191
Patent No. 6376214
GENERAL INFORMATION:
APPLICANT: KUMAR, SANJAY
TITLE OF INVENTION: DRUG BINDING PROTEIN
CORRESPONDENCE ADDRESS: 4
NUMBER OF SEQUENCES: 4
ADDRESSER: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/802,191
FILING DATE: 18-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM T. HAN,
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: GP50000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1310 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-802-191-1
Alignment Scores:
Pred. No.: 0.818 Length: 1310
Score: 77.50 Matches: 39
Percent Similarity: 44.30% Conservative: 27

Best Local Similarity:	26.17%	Mismatches:	50
Query Match:	7.43%	Indels:	33
DB:	4	Gaps:	9

US-10-069-544-2 (1-204) x US-08-802-191-1 (1-1310)

```

Oy 50 LeuAaprglnAlaLeuGlnIleuSerAlaIalaLeuSerAlaIaAaprglnPhnheAlaPro 69
Db 691 ATTGACCAAGCTGAGAGCCGATCATGAAAGTGG-----GGACACACC 732
Oy 70 AlaTyrArgThrLeuAlaLeuValTyrGlnAlaSerGluAapraIaThrInleGlnThrLys 89
Db 733 AGCCCTGAGTCTTCGGCAAAATC-----TCTCTGAAACAAGCC----- 771
Oy 90 AlaGlnAArgLeuPhnGlnLysAlaIaIleGluLysuAenPro---LysAapMerGlnSerTyr 108
Db 772 -----CGAACAATATATCCAGTCCCTGCCCCCATGCCCCAGAAAGAACTTGAGCAAGCATC 825
Oy 109 MetAapTyrGlyPheTyrLeuValGlnMetGlyAapLeuSerGlyAlaLeuIleTyrPhe 128
Db 826 TTCGGTGAGAGCAACCCCTGGGCATA-----GACCTCTTTGAAAGATGCTGGTGGTGC 879
Oy 129 AapLysPProSerAArgAlaIleGlyTyrGluGluArgValAlaIaIleGluLysuMetAla 148
Db 880 GACAGCT-----GACCAAGAGGTCAAGTGCAGTGCAGGCACTGGCC 918
Oy 149 TyrIleTyrTyrHISGlnTyr---GluAlaAlaLysSerProThrLysAapAapTyrAen 167
Db 919 CACGCGTACTTCAAGCAAGTACCAAGCAACCCCGAGAGTAGAGCCAGAGGCGGAGCATATAT 978
Oy 168 AenAlaLysSerAlaLeuGlnAArgAlaLeuAlaLeuIleSerGlyThrGlnInleAapGluIleLys 187
Db 979 GAGAGCGTTGAGGCCCAAGAGCGACGCAAGCTG-----GAGAGGTGGAAG 1020
Oy 188 Lys---SerTyrAapLysLeuLysSer 195
Db 1021 GAGCTCACTTACCGAAAGTCTTAAAGC 1047

```

RESULT 37
116-09-286

```

US-09-286-904-23
Sequence 23, Application US/09286904A
Patent No. 6140124
GENERAL INFORMATION:
APPLICANT: Monia, Brett P.
APPLICANT: Gaarde, William A.
APPLICANT: Nero, Pamela S.
APPLICANT: McKey, Robert
TITLE OF INVENTION: Activase Oligonucleotide Modulation of p38 Mitogen
FILE REFERENCE: ISPH-0347
CURRENT APPLICATION NUMBER: US/09/286,904A
CURRENT FILING DATE: 1999-04-06
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 23
LENGTH: 2180
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (20)..(1138)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: U53442
DATABASE ENTRY DATE: 1996-07-30
US-09-286-904-23

```

Alignment Scores:	
Pred. No.:	18
Score:	77.50
Percent Similarity:	44.30%
Best Local Similarity:	26.17%
Query Match:	7.43%
DB:	3
Length:	2180
Matches:	29
Conservative:	37
Mismatches:	50
Indels:	33
Gaps:	9

US-10-069-544-2 (1-204) X US-09-286-904-23 (1-2180)

[illegible]

RESULT

```

US-09-640-101-23
; Sequence 23, Application US/09640101
; Patent No. 6448079
; GENERAL INFORMATION:
; APPLICANT: Morita, Brett P.
; APPLICANT: Gaarde, William A.
; APPLICANT: Nero, Pamela S.
; APPLICANT: McKay, Robert
; TITLE OF INVENTION: Antisense Modulation of p38 Mitogen
; TITLE OF INVENTION: Activated Protein Kinase Expression
; FILE REFERENCE: ISPN-0468
; CURRENT APPLICATION NUMBER: US/09/640.101
; CURRENT FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 09/266,904
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 2180
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (20)..(1138)
; DATABASE ACCESSION NUMBER: U53442
; DATABASE ENTRY DATE: 1996-07-30
US-09-640-101-23

```

Pred. No.:	18	Length:	2180
Score:	77.50	Matches:	39
Percent Similarity:	44.33%	Conservative:	27
Best Local Similarity:	26.11%	Mismatches:	50
Query Match:	7.43%	Indels:	33
DB:	4	Gaps:	9

US-10-069-544-2 (1-204) X US-09-640-101-23 (1-2180)

QY 50 LeuaprglnalalysglnlnleuapalalaleuseralasparglnphealaPro 69
ATTGACCAAGCTGACGATGAGAGTGTG-----GGCACACCC 769
Db 728 ATTGACCAAGCTGACGATGAGAGTGTG-----GGCACACCC 769
QY 70 AlaTyArgthlreulalalysValTyrglnalasergluaspaalThrhisglnthrlys 89
AGCCCTGAGGTTCTGGCAAAATC-----TCCTGGAAACAGCC----- 808
Db 770 AGCCCTGAGGTTCTGGCAAAATC-----TCCTGGAAACAGCC----- 808
QY 90 AlaGlnAglleuPheglulysalallegluleuAnPro---LysAspMetGlnSerTy 108
CGACATATATCCAGTCCCTGCCCCCAAGAGCCAGACCTGAGCAGATC 862
Db 809 -----CGACATATATCCAGTCCCTGCCCCCAAGAGCCAGACCTGAGCAGATC 862
QY 109 MetAspTyrglyPheTyrglyleuValGlnMetGlyAspLeuSerGlyAlaLeuileTy 128
TTCCCTGAGGACCAACCCCTGCCAT-----GACCTCTTGGAAGATGCTGCTGCTG 916
Db 863 TTCCCTGAGGACCAACCCCTGCCAT-----GACCTCTTGGAAGATGCTGCTGCTG 916
QY 129 AspLysProSerArgAlalleglyTyrglyAlaValAlalleglulysMetAla 148
GACACT-----GACCAAGGGTGCTGAGCTGAGGACTGAGCC 955
Db 917 GACACT-----GACCAAGGGTGCTGAGCTGAGGACTGAGCC 955
QY 149 TyrlleTyTyrgHisglnTyrglnalalalysSerProThrLysAspAspTy 167
CAAGCTCTTACTGACGACATGACAGACCCGAGAGATGAGCCAGAGCCAGCATATGAT 1015
Db 956 CAAGCTCTTACTGACGACATGACAGACCCGAGAGATGAGCCAGAGCCAGCATATGAT 1015
QY 168 AsnalalysSerAlaleuGlnArgAlaleuileSerGlyThrGlnHisAspGlnle 187
GAGACGCTTGAGGCCAAGAGCGCAGCGCTG-----GAGGAGTGGAAG 1057
Db 1016 GAGACGCTTGAGGCCAAGAGCGCAGCGCTG-----GAGGAGTGGAAG 1057
QY 168 Lys---SerTyAspLysLeuLeuSer 195
GAGCTCATTACGAGAGATGCTTAC 1084
Db 1058 GAGCTCATTACGAGAGATGCTTAC 1084
RESULT 39
US-09-134-001C-644
Sequence 644, Application US/09134001C
Patent No. 6180370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 644
LENGTH: 1248
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-644
Alignment Scores:
Pred. No.: 0.887 Length: 1248
Score: 77.00 Matches: 42
Percent Similarity: 38.20% Conservative: 47
Best Local Similarity: 18.03% Mismatches: 100
Query Match: 7.38% Indels: 44
DB: 4 Gaps: 9
US-10-069-544-2 (1-204) x US-09-134-001C-644 (1-1248)
QY 3 IleArgVallyrTriProMetValMetAlaMetGlyLeuValleuSerAlaCysGlnSer 22
ATTATGCTGCTTTACGCAAGTATTAGACCTTGCGCTTATTTCAATAGTGTCTACGCC 150
Db 91 ATTATGCTGCTTTACGCAAGTATTAGACCTTGCGCTTATTTCAATAGTGTCTACGCC 150
QY 23 ThrProIleProProLysAsnAsnProGlnleuAlGlnIleArgthrglnleale 42
GACTCGTTACCAAAAAGATGTGCAAC-----CAAAAAACAATAAGTCACTGTC 204
Db 151 GACTCGTTACCAAAAAGATGTGCAAC-----CAAAAAACAATAAGTCACTGTC 204

QY 43 SerleuLeuasp-----MetGlyLysleuaspGlnalalalysglnlnleuasp 58
AGTATATAAGCGTTCCAGATGCGCAAACTTCTGTAAGACATATTATCTCGT 264
Db 205 AGTATATAAGCGTTCCAGATGCGCAAACTTCTGTAAGACATATTATCTCGT 264
QY 59 AlaAlaleu-----SerAlaaspArglnPheAlaProAlaTyArgthlreula 75
CTACCTTATTAGTAAAGCTTCCAACCAACCAAGCAACATCGTAT---ACACTGTGAA 321
Db 265 CTACCTTATTAGTAAAGCTTCCAACCAACCAAGCAACATCGTAT---ACACTGTGAA 321
QY 76 ---LysValTyrglnalasergluaspaal-----ThHis 86
CCTTTTAAATTTAATTTAATTAAGAAAGCAGCGCAATTATTAATTCAGTCTC 381
Db 322 CCTTTTAAATTTAATTTAATTAAGAAAGCAGCGCAATTATTAATTCAGTCTC 381
QY 87 GlnThrLysAlaGlnArgleuPheglulysalallegluleuAnProLysAspMetGln 106
AATAAAGAGATGTCATTTATGTATGATACAAATTTCTCTAATCTTCAAAATTCATA 441
Db 382 AATAAAGAGATGTCATTTATGTATGATACAAATTTCTCTAATCTTCAAAATTCATA 441
QY 107 SerTyMetAspTyrglyPheTyrglyleuValGlnMetGlyAspLeuSerGlyAlaLeuile 126
GCTTCAAAACAGCAAAACATATTATTCATTAATGTTTCTCATTTCTTCAAAATATA 501
Db 442 GCTTCAAAACAGCAAAACATATTATTCATTAATGTTTCTCATTTCTTCAAAATATA 501
QY 127 -----TyPheAspLys 130
AATCAATATATAAAATCAAAAGATTAACAATTGACTAATACAAAAGATATT 555
Db 502 AATCAATATATAAAATCAAAAGATTAACAATTGACTAATACAAAAGATATT 555
QY 131 ProSerArgAlalleglyTyrglyAlaValAlaile-----GluAsn 146
GCACTTACTGAAGATGTAAAGTGAACCTTGCTGCTTAAACCCAGCTAT 606
Db 556 GCACTTACTGAAGATGTAAAGTGAACCTTGCTGCTTAAACCCAGCTAT 606
QY 147 MetAlaTyrlleTyTyrgHisglnTyrglnalalalysSerProThrLysAspAspTy 166
AATGAAAAACATATGAATAATGACCATGAATCCATTAAGATTTAAAGATTTT 666
Db 607 AATGAAAAACATATGAATAATGACCATGAATCCATTAAGATTTAAAGATTTT 666
QY 167 AsnAsnalalysSerAlaleuGlnArgAlaleuileSerGlyThrGlnHisAspGlnle 186
AAACAACTGCATGATTAACAACCAACTTGAATATCAAGATCAACGAATGAATG 726
Db 667 AAACAACTGCATGATTAACAACCAACTTGAATATCAAGATCAACGAATGAATG 726
QY 187 LysLysSerTyAspLysLeuLeuSerAspTyTy 199
TACGAGATATGTAATCAATTAAGATTTCAAGATA 765
Db 727 TACGAGATATGTAATCAATTAAGATTTCAAGATA 765
RESULT 40
US-08-687-956A-22
Sequence 22, Application US/08687956A
Patent No. 5861157
GENERAL INFORMATION:
APPLICANT: BURNIE, JAMES P
APPLICANT: MATHEWS, RUTH C
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
TITLE OF INVENTION: INFECTIONS DUE TO STREPTOCOCCI AND ENTEROCOCCI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO, LLP
STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH
FLOOR
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,956A
FILING DATE: 29-JUL-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9401689.6
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N
REGISTRATION NUMBER: 16,773

Job time : 121 secs

REFERENCE/DOCKET NUMBER: 50885/222892
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/861-3000
TELEFAX: 202/822-0944
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 5077 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus sobrinus
US-08-687-956A-22

Alignment Scores:

Pred. No.:	7.84	Length:	5077
Score:	77.00	Matches:	55
Percent Similarity:	42.06%	Conservative:	35
Best Local Similarity:	25.70%	Mismatches:	76
Query Match:	7.38%	Indels:	49
DB:	2	Gaps:	12

US-10-069-544-2 (1-204) x US-08-687-956A-22 (1-5077)

QY 21 GlnSerThrProIleProProlYsAsnProGlnLeuAlaGlnIleArgThrGlnIle 40
DB 1038 CAGAAACGCGCTATCCAGCAACCAATGCCAAGCTAAGGCAATTACGAAACCA--- 1094
QY 41 AlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAla 60
DB 1095 -----ACTGGCTCAATATGATAAAGATTTA-GCCGACGCCCACTGCTAATGCTACA 1147
QY 61 LeuSerAlaAspArgGlnPheAlaProAla-----TyrArgThrLeuAlaLysValTyr 78
DB 1148 AATGAGCGGAGCTACCAAGCTAAGAGCAGCTTATGAACAAGATTAGCGCGCTGCAA 1207
QY 79 GlnAlaSerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIle 98
DB 1208 GCCGCTAAT-----GCAGCTGCCAAGCAGCGCTTACGAACAAAGCTCTA 1249
QY 99 GluLeuAsn-----ProLys 103
DB 1250 GCTGCCAACAGCGCCAGAACGCCAATCAAGCCGGAATAAGAGCTATCCAGCAGCGC 1309
QY 104 AspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGln--MetGlyAspLeuSer 122
DB 1310 AATGCGCAAGCTAAGCTAATGATGAATTAATGACCAATATCAAAAGGATTGGCC 1369
QY 123 GlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArgValVal 142
DB 1370 GCAGCT-----CAATCTGTGTAACCGCTAATGAGCAGACTTACCAAGAAAATTAGCA 1423
QY 143 AlaIleGlu---AsnMetAlaTyrIleTyrTyrHisGlnTyrGluAlaAlaLysSerPro 161
DB 1424 GCCTATGAAAGAAAGAACTGGCTGCTG-----CAAGCAGCCCAATGCACGT 1468
QY 162 ThrLysAspAspTyrAsn-----AsnAlaLysSerAla----- 172
DB 1469 GCTAAGCAAGATATGACAGAAAGTTTCAGAGACTAATGCTAATAAATGCCAATATACG 1528
QY 173 ---LeuGluArgAlaLeuIleSerGlyThrGlnHisAspGluIleLysLysSerTyrAsp 191
DB 1529 GAAGCAACCGGCTATC-----CGTAACGCAATGCCAAGCCCAAGACAGACTATGAA 1582
QY 192 LysLeuLeuSerAspTyr---LysLeuLeuSerAspTyrLys 204
DB 1583 CTCAAACTGCTTAAGTACCAAGAGAGCTTGTCTCACTACAG 1624

Search completed: July 4, 2003, 04:06:02

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 4, 2003, 03:57:48 ; Search time 123 Seconds
(without alignments)
2568.961 Million cell updates/sec

Title: US-10-069-544-2
Perfect score: 1043
Sequence: 1 MKIRVKNPMWAMGLVLSAC.....EIKSYDKLSLDYKLSLDYK 204

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 6.0, Ygapext 7.0
Delop 6.0, Delext 7.0

Searched: 1085931 seqs, 780495707 residues
Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPRO.spool/US10069544/runat_30062003.091540.26377/app.query.fasta.1.391
-DB=Published Applications_NA -OPMT=faetap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bite -START=1 -END=1 -MATRIX=bl0sum62
-TRANS=trans40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=0 -MODE=LOCAL -OUTPMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10069544.@CGN.1.1.118 @runat_30062003.091540.26377
-NCPU=6 -ICU=3 -NO MMAP -LARGEBUERT -NEG_SCORES=0 -WAIT -DSPLOCK=100
-LONLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:
1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubna/PC1_NEM_PUB.seq:
3: /cgn2_6/ptodata/1/pubna/US06_NEM_PUB.seq:
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubna/US07_NEM_PUB.seq:
6: /cgn2_6/ptodata/1/pubna/PC1US_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubna/US08_NEM_PUB.seq:
8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubna/US09_NEM_PUB.seq:
10: /cgn2_6/ptodata/1/pubna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubna/US10_NEM_PUB.seq:
12: /cgn2_6/ptodata/1/pubna/US10_PUBCOMB.seq:
13: /cgn2_6/ptodata/1/pubna/US60_NEM_PUB.seq:
14: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	118.5	11.4	1830121	9	US-10-329-960-1	Sequence 1, Appli
2	108	10.4	2010	9	US-10-156-761-6557	Sequence 6557, Ap
3	108	10.4	9025608	9	US-10-156-761-1	Sequence 1, Appli
4	106	10.2	1971	10	US-09-974-300-1935	Sequence 1935, Ap

Result No.	Score	Match	Length	DB	ID	Description
5	98.5	9.4	3084	10	US-09-880-107-3426	Sequence 3426, Ap
6	98.5	9.4	3586	9	US-09-919-039-75	Sequence 75, Appl
7	98.5	9.4	4799	10	US-09-925-297-214	Sequence 214, App
8	98.5	9.4	6812	9	US-10-158-646-29	Sequence 29, App
9	98.5	9.4	1241	9	US-10-158-646-12376	Sequence 12376, A
10	95	9.1	6465	10	US-09-954-456-193	Sequence 193, App
11	92.5	8.9	3080	9	US-10-197-666A-153	Sequence 153, App
12	92.5	8.9	3253	9	US-10-197-666A-151	Sequence 151, App
13	92	8.8	1994	9	US-10-108-605-38	Sequence 38, Appl
14	92	8.8	2909	9	US-10-197-666A-121	Sequence 121, App
15	90	8.6	1632	10	US-09-917-800A-1526	Sequence 1526, Ap
16	90	8.6	1824	12	US-10-044-090-543	Sequence 543, App
17	90	8.6	1866	12	US-10-044-090-542	Sequence 542, App
18	90	8.6	2230	9	US-10-198-846-14067	Sequence 14067, A
19	90	8.6	3758	9	US-10-193-950A-1	Sequence 1, Appli
20	89	8.5	3309400	9	US-09-738-626-1	Sequence 1, Appli
21	88.5	8.5	13086	7	US-08-781-986A-16	Sequence 16, Appl
22	88	8.4	621	10	US-09-974-300-2413	Sequence 2413, Ap
23	87.5	8.4	1317	9	US-10-043-487-174	Sequence 174, App
24	87	8.3	31880	9	US-10-114-170-242	Sequence 242, App
25	86	8.2	2031	9	US-10-128-714-2166	Sequence 2166, Ap
26	86	8.2	2031	9	US-10-128-714-2166	Sequence 2166, Ap
27	86	8.2	2082	9	US-10-128-714-2166	Sequence 2166, Ap
28	86	8.2	2082	9	US-10-128-714-2166	Sequence 2166, Ap
29	86	8.2	2082	9	US-10-128-714-2166	Sequence 2166, Ap
30	86	8.2	4082	9	US-10-128-714-2166	Sequence 2166, Ap
31	85.5	8.2	496	9	US-10-198-846-9055	Sequence 9055, Ap
32	85.5	8.2	1128	9	US-09-938-842A-2043	Sequence 2043, Ap
33	84.5	8.1	2113	9	US-09-880-107-2411	Sequence 2411, Ap
34	84.5	8.1	4677	9	US-09-938-842A-5	Sequence 5, Appli
35	84.5	8.1	6225	9	US-09-978-244A-7	Sequence 7, Appli
36	84	8.1	2613	9	US-09-981-353-163	Sequence 163, App
37	84	8.1	2613	9	US-10-158-646-27	Sequence 27, Appl
38	84	8.1	3540	10	US-09-815-242-9510	Sequence 9510, Ap
39	83.5	8.0	1783	9	US-10-106-698-2038	Sequence 2038, Ap
40	83	8.0	1731	9	US-10-153-668-193	Sequence 193, App
41	83	8.0	2901	10	US-09-801-368-371	Sequence 371, App
42	83	8.0	4017	9	US-10-153-668-195	Sequence 195, App
43	82.5	7.9	2362	9	US-09-919-039-364	Sequence 364, App
44	82	7.9	1641	9	US-10-128-714-2384	Sequence 2384, Ap
45	82	7.9	1884	9	US-09-938-842A-220	Sequence 220, App

ALIGNMENTS

RESULT 1
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329, 960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643, 990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487, 429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426, 787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:

```
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)

LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100691)..(100691)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
```

```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

```

```

Alignment Scores:
Pred. No.: 0.268 Length: 1830121
Score: 118.50 Matches: 39
Percent Similarity: 39.47% Conservative: 21
Best Local Similarity: 25.66% Mismatches: 89
Query Match: 11.36% Indels: 3
DB: 9 Gaps: 1

```

US-10-069-544-2 (1-204) x US-10-329-960-1 (1-1830121)

```

Qy 1 Metyleleargvalivtrpmetvalmetalegluvalleuseralacys 20
Db 389428 ATGAAACAACTAGTAACTAAGTGGGTATTTCCCTTCATTTTCTGCGC 389487
Qy 21 Glnserthrproleproprolyasansproglinleaglileargthrlnle 40
Db 389488 GTTTCACATGTCATCTAGCTTAATCATCAATGCCGCTAAAGCCGCTGGAATC 389547
Qy 41 Alalileserleuaspmetglylyleuaspglinalalyglnleuaspalaia 60
Db 389548 GCTTGAAGCTATCTTCAGCAAAATATATCTCACTGCGCTAAATCATTTAGACAAAGCA 389607
Qy 61 leuseralasparglnphealaproalaityrargthrleualalyvaltyrglnla 80
Db 389608 CTTCAACACGATTAATAATTAATCTGCTGATTCAGCACTTGCACTTATATCAACAA 389667
Qy 81 sergluapalathrhieglnthryleaglinalrleupheglulyalailegluleu 100
Db 389668 CAAGGCGAAATAGAAAT-----GCTTTCGTGAATATGAATAGCCGTAAACCTT 389718
Qy 101 leuprolyasapmetglnsertrymetaspyrlyrphetyrleuvalglnmetelyasp 120
Db 389719 AATCATTAACAAGGCGATGTCATTAATTTGGTACGTTTCATGATGCAAAAGAAA 389778
Qy 121 leuserglylaleuiletyrphesplysproseraglaileglytyrgluqlarg 140
Db 389779 TTGGAACAAGCTCAGCAACATTTGAATTAGACACTTAATTCGCGAATATTATCATCA 389838
Qy 141 Valvalalalegluasmetalatyrlletyrtyr 152
Db 389839 GCAAGTACATTTGAAATATATCGTCTTGCTGCTTAT 389874

```

RESULT 2
US-10-156-761-6557

```

Sequence 6557, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASHIYA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272897
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 6557
LENGTH: 2010
TYPE: DNA
ORGANISM: Streptomyces avermectilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2010)
US-10-156-761-6557

```

```

Alignment Scores:
Pred. No.: 0.000235 Length: 2010
Score: 108.00 Matches: 47
Percent Similarity: 36.67% Conservative: 30
Best Local Similarity: 22.38% Mismatches: 103
Query Match: 10.35% Indels: 31
DB: 9 Gaps: 3

```

US-10-069-544-2 (1-204) x US-10-156-761-6557 (1-2010)

```

Qy 3 Ileargvalivtrpmetvalmetalegluvalleuseralacysglnser 22
Db 719 GTGCGCGCGGTGGCGGTGCGACTACCGCGGGCGGTGCTGCGCGCTCCGG 778
Qy 23 Thrproleproprolyasansproglinleaglileargthrlnlealaile 42
Db 779 ACGCGGAACCGCAAGCCGCGCGCGCG-----CGTGGCC-GTCTCCCG 825
Qy 43 Serleuaspmetglylyleuaspglinalalyglnleuaspalaialeuser 62
Db 826 CAATGCTGCCAGTGGCGAGCTGACGCGAGCATCACTCCCTCCAGGCTCACCTCCG 885
Qy 63 Alasparglnphealaproalaityrargthrleualalyvaltyr----- 78
Db 886 GCCAGCCCAAGACTTGCGCGGTGGCGAGCGCTCGCGCTCGGTATCATGACAGGCC 945
Qy 79 Glnalasergluapalathrhieglnthryleaglinalrleupheglulyalaile 98
Db 946 CGACCAAGGGGCAACCCCTCCGCTACCGCGAGCGAGCGGACGACGACCTCCCTG 1005
Qy 99 GluleuaspProlyasapmetglnsertrymetaspyrlyrphetyrleuvalglnmet 118
Db 1006 CGACTGCGCGCGCAAGCAAGCCGCGCGCTGCGAGCGCGCGCTCGCGCGCGCG 1065
Qy 119 Glyaspsergluvalaleuiletyrphesplysproseraglaileglytyrglu 138
Db 1066 CACGACTTCCAGCGCGCTGACCTACGAGCGCGCGCTGAAGCAAAACCGTACAGC 1125
Qy 139 Glyargvalalalegluasmetalatyrlletyrtyrhieglnlyrglualaia 158
Db 1126 GAGCGCGCGCTGCTCCGATGACAGCGCTGTCGAACTCGCGCGCTGACGAGCGG 1185
Qy 159 LyaserProthr----- 162
Db 1186 TCGAAGCGCGCGAGCTGCGCGAGCAAGCGCGCGCGCGCTCCGATTCACCGCGTAC 1245

```

Qy 163 -----LysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAla 176
 Db 1246 GCCTACGTCGGGGAAGCTCGGGGGGAGCTGAACCCGCCCGCTCTGAGAGAGCG 1305
 Qy 177 LeuIleSerGlyThrGlnHisAspGluIle 186
 Db 1306 CTGAGACCGCGCAGCGCGCGCGGAGCATC 1335

RESULT 3

US-10-156-761-1/c
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
80.7	108.00	9025608	47
Percent Similarity:	36.67%	Conservative:	30
Best Local Similarity:	22.38%	Mismatches:	103
Query Match:	10.35%	Indels:	31
DB:	9	Gaps:	3

US-10-069-544-2 (1-204) x US-10-156-761-1 (1-9025608)

Qy 3 11aTgVAllyETpRomeValMeAlaMetGlyLeuValLeuSerAlaCySgInser 22
 Db 7877206 GTGCGGCGCGTGGGCGGTGCGGACTCACCGCGGCGCGTGGTGGCGCGCGG 7877147
 Qy 23 ThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIle 42
 Db 7877146 AGCGGCGGAACCGCATATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7877100
 Qy 43 SerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSer 62
 Db 7877099 CAATGCTCGCGAGTGGCGACCTGGACGAGAGATCACTCCCTCCAGGCTCACCTCCGG 7877040
 Qy 63 AlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyr----- 78
 Db 7877039 GCCCGACCCAGAGACTTCGCGCGGTGGCGAGCGCTCGCGCGGTACATCGACGAGCGC 7876980
 Qy 79 GlnAlaSerGlnAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIle 98
 Db 7876979 CGCACCAGAGGCGAGCCCTCCGCTACCGCGAGCGCAGCGGCGACTGAGACGCTCCG 7876920
 Qy 99 GluLeuAspProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMet 118
 Db 7876919 CGACTGCGCGCGAGCAACGACCGCGGCGTGGCGGAGCGCGCGCGCTCCCGCGCGCGCG 7876860

Qy 119 GlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGlu 138
 Db 7876859 CACACTCTTCACCGCGCGCTGACCTTACCGCGAGCGCGCTGAACACCGCTACAGC 7876800
 Qy 139 GlyArgValAlaIleIleGluAsnMetAlaTyrIleTyrThrHisGlnTyrGluAlaIle 158
 Db 7876799 GAGTGGCCCTGTCCTCCGTATCGAGCCCTGCTGCAACTCGCGCGCTACGAGAGCG 7876740
 Qy 159 LysSerProThr----- 162
 Db 7876739 TCGAAGCGCGCGACGTGGCGCGACCAAGCGCGCGCTCCGCTTTCACGCGGTAC 7876680
 Qy 163 -----LysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAla 176
 Db 7876679 GCCTACGTCGGGGAAGCTCGGGGGGAGCTGAACCCGCCCGCTCTGAGAGAGCG 7876620

RESULT 4

US-09-974-300-1935
 ; Sequence 1935, Application US/09974300
 ; Parent No. US20020146721A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Berka, Randy M.
 ; APPLICANT: Clausen, Ib Groch
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
 ; FILE REFERENCE: 10085,500-US
 ; CURRENT APPLICATION NUMBER: US/09/974,300
 ; CURRENT FILING DATE: 2001-10-05
 ; PRIOR APPLICATION NUMBER: 09/680,598
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: 60/279,526
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 8481
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1935
 ; LENGTH: 1971
 ; TYPE: DNA
 ; ORGANISM: Bacillus licheniformis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)....(1971)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-974-300-1935

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0.000428	106.00	1971	47
Percent Similarity:	41.71%	Conservative:	31
Best Local Similarity:	25.13%	Mismatches:	73
Query Match:	10.16%	Indels:	36
DB:	10	Gaps:	8

US-10-069-544-2 (1-204) x US-09-974-300-1935 (1-1971)

Qy 18 SerAlaCySgInserThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArg 37
 Db 219 ACACTGCGCGGTGCGGTCCCGAGCGGTGAACCA-----CAGCTGATCAAGTCCCG 272
 Qy 38 ThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeu 57
 Db 273 AAGACATCGGCATCGAACATTAACAGCAGCAAAAGTATCGGCTGCTTGAACAA-- 329
 Qy 58 AspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysVal 77
 Db 330 ---GCGGGATGGGACACGCTTCATCAAGCGCGCGCAAAACGTTACATGTAATAATTG 386
 Qy 78 TyrGlnAlaSerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGlnLysAla 97
 Db 7876919 CGACTGCGCGCGAGCAACGACCGCGGCGTGGCGGAGCGCGCGCGCTCCCGCGCGCGCG 7876860

Db 387 AAGAGGCAAAACCGGATCTTTCAATCTAAAGCCCATCATCCGGCTCCAGCA 446
Qy 98 IleguleAenProlysaPwMetGlnserTyrmecAspTyrglyPheTyrlenuValgln 117
Db 447 TTAGAGCCCTCGGTAAAGCAAAACGGCATTAAAGAT-----ATGTACCTTGGAA 500
Qy 118 MetGlyAspLeuSerGlyAlaLeuIleTyrlPheAspLysProSerArgAlaIleGly--- 136
Db 501 ATG-----MAGAAAGCGTCTTCGCGATTC 527
Qy 137 -----TyrluclYarValValAlaIleGluAsnMetAlaTyrlle 150
Db 538 AAGGCGCGCAAGCAATATCAGCGCGCCTCCAGATGCCAAGATGCTGAATG 587
Qy 151 TyrlYr-----HisGlnTyrgluAlaAlaLysSerProThrLysAspTyrl 166
Db 588 GAATTCGTCGAACAGGTTCAAGCATATGAGCATCTCAAGAGACCCCAATATCATAT 647
Qy 167 -----AsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGlyThr 181
Db 648 CAATGTCGCTCAAAAATAACGCAAAAC-----CGTGTGCGAATGCTGACATA 698
Qy 182 GlnHisAspGluIleLysLys 188
Db 699 CAATTGAACAGCTTCAGAAAG 719

RESULT 5

US-09-880-107-3426
; Sequence 3426, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Scherf, Joseph G.
; APPLICANT: Scherf, Joseph G.
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT FILING DATE: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3426
; LENGTH: 3084
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U77413
US-09-880-107-3426

Alignment Scores:

Pred. No.: 0.00884 Length: 3084
Score: 98.50 Matches: 52
Percent Similarity: 40.31% Conservative: 27
Best Local Similarity: 26.53% Mismatches: 82
Query Match: 9.44% Indels: 36
DB: 10 Gaps: 8

US-10-069-544-2 (1-204) x US-09-880-107-3426 (1-3084)

Qy 17 LeuSerAlaCyGlnSerThrProIleProProLyAsnAsnProGlnLeuAlaGlnIle 36
Db 843 CTCCTCCGTCTGTGT-CCCAACCATGACACTCTCTGAATAAC-----CTAGCCCAATATC 895
Qy 37 ArgThrGlnIleAlaIleSerLeuAspMetGlyLysLeuAspGlnAlaLysGlnGln 56
Db 896 AAACGA-----GAACGAGAAACATTGAAGAGGCACTTGCCTTG 934
Qy 57 LeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrlArgThrLeuAlaLys 76

Db 935 TATGTAAAGCATTAAGAGTCTTCCAGAGTTTCGTCTGCCCATTCGAATTTAGCAAGT 994
Qy 77 ValTyrglnAlaSerGluAspAlaThrHisGlnThrLysAlaGlnArgLeu----- 93
Db 995 GTACTGCAG-----CAGCAGGAAACATGCAAGAGCTTGATGCAT 1036
Qy 94 PheGluLysAlaIleGluLeuAsnProLysAspMetGlnserTyrmecAspTyrglyPhe 113
Db 1037 TATAAGAGGCTATTTCGAATACGTCTACCTTTCGTATGCTCTCTCAATATAGGAAC 1096
Qy 114 TyrlenuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrlPheAspLysProSerArg 133
Db 1097 ACTCTTAAGAGAGATGACGAGATGTTTCAGAGAGCTTGAGATGTAT-----ACCGCT 1147
Qy 134 AlaIleGlyTyrgluGluArgValVal-----AlaIleGluAsnMetAlaTyrlleTyrl 151
Db 1148 GCATTCGAATTAATCTCTGATTCGATTCGACATGACATGCAATTCGGCTTCATTCATTAAG 1207
Qy 152 -----TyrlHisGlnTyrgluAlaAlaLysSer 160
Db 1208 GATTCAGGAATATTCAGAAAGCATAGCTTCTTACCGCAGCGCTCGAAATTAAGCT 1267
Qy 161 ProThrLysAspAspTyrlAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
Db 1268 GATTTCTCGATGCTTATTTGAATCTTGATGCTCATTCGCTGAG--ATTGTCTGTGATTGG 1324
Qy 181 ThrGlnHisAspGluIleLysLysSerTyrlAspLysLeuSerAlaP 196
Db 1325 ACAGACTATGATGAGCGAATGAAGAGTGTGTCAGTATTTGTGCTGAC 1372

RESULT 6

US-09-919-039-75
; Sequence 75, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT FILING DATE: US/09/919,039
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 75
; LENGTH: 3586
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 3766715CBI
US-09-919-039-75

Alignment Scores:

Pred. No.: 0.0111 Length: 3586
Score: 98.50 Matches: 52
Percent Similarity: 40.31% Conservative: 27
Best Local Similarity: 26.53% Mismatches: 82
Query Match: 9.44% Indels: 36
DB: 9 Gaps: 8

US-10-069-544-2 (1-204) x US-09-919-039-75 (1-3586)

Qy 17 LeuSerAlaCyGlnSerThrProIleProProLyAsnAsnProGlnLeuAlaGlnIle 36
Db 849 CTCCTCCGTCTGTGT-CCCAACCATGACACTCTCTGAATAAC-----CTAGCCCAATATC 901
Qy 37 ArgThrGlnIleAlaIleSerLeuAspMetGlyLysLeuAspGlnAlaLysGlnGln 56
Db 902 AAACGA-----GAACGAGAAACATTGAAGAGGCACTTGCCTTG 940
Qy 57 LeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrlArgThrLeuAlaLys 76

Db 941 TATCGTAAGCATTTAGAACTCTTCCAGAGCTTTCCTGCCCATTTCAATTAGCACT 1000
Qy 77 ValTyrGlnAlaSerGluAspAlaThrHisGlnThrValAlaGlnArgLeu----- 93
Db 1001 GTACGCGAG-----CAGCAGGGGAAACTGCAGAGACTCTGATGCAT 1042
Qy 94 PheGluValAlaAlaGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPhe 113
Db 1043 TATAGAGAGCTATGATGCAATCGACTTCTTACCTTGCTGATGCTACTTAAATGGAAC 1102
Qy 114 TyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArg 133
Db 1103 ACTCTTAAGAGATGAGATGATCTTCAGAGACTTCGACTGTAT-----ACGGCT 1153
Qy 134 AlaIleGlyTyrGluGlyArgValVal-----AlaIleGluAsnMetAlaTyrIleTyr--- 151
Db 1154 GCCATCAAAATTAACTCGCATTTGCAGATGCACATAGCAATCTGGCTTCATTCTAAG 1213
Qy 152 -----TyrHisGlnTyrGluAlaAlaLysSer 160
Db 1214 GATTCAGGAATATTCAGAGCCATAGCTTCTTACCGCAGCTCTGAACCTTAAGCT 1273
Qy 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
Db 1274 GATTTCTGATGCTTATGTAATCTTGCTCATGCTGCTGAG---ATTGCTGTGATTGG 1330
Qy 181 ThrGlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSerAsp 196
Db 1331 ACAGACTATGATGAGCGAATGAAGAAGTTGTCAGATTGCTGCTGAC 1378

RESULT 7

US-09-925-297-214
Sequence 214, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 214
LENGTH: 4799
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1164)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (2403)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-214

Alignment Scores:
Pred. No.: 0 0173 Length: 4799
Score: 98.50 Matches: 52.
Percent Similarity: 40.31% Conservative: 27
Best Local Similarity: 26.53% Mismatches: 82
Query Match: 9.44% Indels: 36
DB: 10 Gaps: 8

US-10-069-544-2 (1-204) x US-09-925-297-214 (1-4799)

Qy 17 LeuSerAlaCysGlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIle 36
Db 470 CTCCTCGCTGTGT-CCACCCATGCAAGCTCTGTAATAC-----CTAGCATATATC 522

Qy 37 ArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGln 56
Db 523 AAACGA-----GACAGGGGAAACATTGAAGCGCTTCGCTTG 561
Qy 57 LeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLys 76
Db 562 TATGTAAGCATTTAGAACTTCCAGAGCTTCTGCTGCCATTCAAATTAGCAAGT 621
Qy 77 ValTyrGlnAlaSerGluAspAlaThrHisGlnThrValAlaGlnArgLeu----- 93
Db 622 GTACGCGAG-----CAGCAGGGGAAACTGCAGAGACTCTGATGCAT 663
Qy 94 PheGluValAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPhe 113
Db 664 TATAGAGAGCTATTCGATGATGCTTCTTACCTTGCTGATGCTCTTAATGGAAC 723
Qy 114 TyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArg 133
Db 724 ACTCTTAAGAGATGAGATGATCTTCAGAGACTTCGACTGTAT-----ACGGCT 774
Qy 134 AlaIleGlyTyrGluGlyArgValVal-----AlaIleGluAsnMetAlaTyrIleTyr--- 151
Db 775 GCCATCAAAATTAACTCGCATTTGCAGATGCACATAGCAATCTGGCTTCATTCTAAG 834
Qy 152 -----TyrHisGlnTyrGluAlaAlaLysSer 160
Db 835 GATTCAGGAATATTCAGAGCCATAGCTTCTTACCGCAGCTCTGAACCTTAAGCT 894
Qy 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
Db 895 GATTTCTGATGCTTATGTAATCTTGCTCATGCTGCTGAG---ATTGCTGTGATTGG 951
Qy 181 ThrGlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSerAsp 196
Db 952 ACAGACTATGATGAGCGAATGAAGAAGTTGTCAGATTGCTGCTGAC 999

RESULT 8

US-10-158-646-29/C
Sequence 29, Application US/10158646
Publication No. US20030073105A1
GENERAL INFORMATION:
APPLICANT: Lasek, Amy K.W.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0030-1 US
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: 60/295,239
PRIOR FILING DATE: 2001-05-31
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PERL Program
SEQ ID NO 29
LENGTH: 6812
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030073105A1 335916.17c
US-10-158-646-29

Alignment Scores:
Pred. No.: 0 0294 Length: 6812
Score: 98.50 Matches: 52
Percent Similarity: 40.31% Conservative: 27
Best Local Similarity: 26.53% Mismatches: 82
Query Match: 9.44% Indels: 36
DB: 9 Gaps: 8

US-10-069-544-2 (1-204) x US-10-158-646-29 (1-6812)

QY 17 LeuSerAlaCyGlnSerThrProIleProProlYAsnAspProGlnLeuAlaGlnIle 36
 |||||
 DB 4239 CTCCTCGCTGTGT-CCCAACCATGAGACTCTCGAATAAC-----CTACCAATATTC 4247
 QY 37 ArgThrGlnIleAlaIleSerLeuMetGlyLysLeuAspGlnAlaLysGlnGln 56
 :||:|
 DB 4246 AAGCA-----GAAACAGGAAACATTGAGAGAGCACTGCTTG 4208
 QY 57 LeuAspAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLys 76
 |||||
 DB 4207 TATCTAAAGCATTTGAAAGCTTTCCAGATTGCTGCTGCTGCTCAATTACAACTTACAGACT 4148
 QY 77 ValTyrGlnAlaSerGlyAspAlaThrIleGlnThrLysAlaGlnArgLeu----- 93
 |||||
 DB 4147 GATACGACG-----CAGCAGGAGAAATGCAAGAGAGCTGATGCAT 4106
 QY 94 PheGlyLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPhe 113
 :||:|
 DB 4105 TATAAGAGGCTATTGCAATGCTTACCTTGTGCTGATCCTACTTAAATATGGAAC 4046
 QY 114 TyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArg 133
 |||||
 DB 4045 ACTTAAAGAGATGCAAGATGCTTCAAGAGCTTGCAGCTGTAT-----ACGGCT 3995
 QY 134 AlaIleGlyTyrGlnGlyArgValAla-----AlaIleGluAsnMetAlaTyrIleTyr--- 151
 |||||
 DB 3994 GCCATCAAAATTAATCCTGCAATTCGACATGCAATGCAATGCAATGCTTCCATTCATAG 3935
 QY 152 -----TyrHseGlnTyrGlnAlaAlaLysSer 160
 DB 3934 GATTCAGGAAATATTCAGAAACCATAGCTTCTTACCGCAGCGCTCTGAAACTTAAGCT 3875
 QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
 |||||
 DB 3874 GATTTTCTGATGCTTATGTAAGCTTGCCTCATTCCTGACG-----ATTGCTGTATGG 3818
 QY 181 ThrGlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSerAsp 196
 |||||
 DB 3817 ACAGCATATGATGAGCAATGAAGAAGTTGCTGATATTTGCGCGAC 3770
 DB
 RESULT 9
 US-10-198-846-12376
 ; Sequence 12376, Application US/10198846
 ; Publication No. US2003009974A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lillie, James
 ; APPLICANT: Xu, Yongyao
 ; APPLICANT: Wang, Youzhen
 ; APPLICANT: Steinmann, Kathleen
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
 ; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 ; TITLE OF INVENTION: THERAPY OF BREAST CANCER
 ; FILE REFERENCE: MRI-049
 ; CURRENT APPLICATION NUMBER: US/10/198,846
 ; CURRENT FILING DATE: 2002-07-18
 ; PRIOR APPLICATION NUMBER: 60/306,220
 ; PRIOR FILING DATE: 2001-07-18
 ; NUMBER OF SEQ ID NOS: 14084
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12376
 ; LENGTH: 1241
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: m1sc_feature
 ; LOCATION: 1, 2, 3, 1239, 1240, 1241
 ; OTHER INFORMATION: n = A,T,C or G
 US-10-198-846-12376
 Alignment Scores:
 Pred. No.: 0.0026 Length: 1241
 Score: 98.00 Matches: 43

Percent Similarity: 39.56%
 Best Local Similarity: 23.63%
 Query Match: 9.40%
 DB: 9 Gaps: 6
 US-10-069-544-2 (1-204) x US-10-198-846-12376 (1-1241)
 QY 31 ProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeu 50
 |||||
 DB 126 CTTCTCAGAGAGAGCTCCGCTTGACAGCTGCTCAGGTTTGGCCGTATGGGTACAGCA 185
 QY 51 AspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAla 70
 :||:|
 DB 186 AAGAGCTGAAAGAGACCAATCACTTGTGCAGAGAGACCGGATGCTTGAATGC 245
 QY 71 TyrArgThrLeuAlaLysValTyrGlnAlaSerLysAspAlaThrIleGlnThrLysAla 90
 |||||
 DB 246 TATGCTCTTGTACGCTTATATGCAAGCAGAG-----AACACAGACAGCA 296
 QY 91 GlnArgLeuPheGlyLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAsp 110
 |||||
 DB 297 CTGATGCTATAGCAAGCTCTCCAGCTGAACCAAGACCCAAAGTATTTCTGA 356
 QY 111 TyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLys 130
 |||||
 DB 357 CTTTTTTT-----ACAAAGAAACCAATTAAGAGAGCAAGACCTTCTGACAAA 407
 QY 131 ProSerArgAlaIleGlyTyrGlnGlyArgValAlaIleGlu----- 145
 |||||
 DB 408 -----GCTTTGAGAGCTATAGAGTGTGTGCAACTTAACCCAGACCA 452
 QY 146 -----AsnMetAlaTyrIleTyrThrIleGlnTyrGlnAlaAlaLysSer 160
 |||||
 DB 453 GCACAGCGCTGATGACATGCGTGCATCAACAC----- 488
 QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
 |||||
 DB 489 ---ATCAAGGAGAAATATGTGTGCAAGAGCTTATATGAGAGAGCTTA----- 536
 QY 181 ThrGlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSerAspTyrLysLeu 200
 |||||
 DB 537 -----CAGCTGTTCCAGACAGCAACTGCTG 563
 QY 201 SerAsp 202
 DB 564 AAGAA 569
 DB
 RESULT 10
 US-09-954-456-193
 ; Sequence 193, Application US/09954456
 ; Patent No. US20020115057A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Young, Paul
 ; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
 ; TITLE OF INVENTION: Sets
 ; FILE REFERENCE: 689290-76
 ; CURRENT APPLICATION NUMBER: US/09/954,456
 ; CURRENT FILING DATE: 2001-09-18
 ; PRIOR APPLICATION NUMBER: US/60/233,617
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: US/60/234,052
 ; PRIOR FILING DATE: 2000-09-20
 ; PRIOR APPLICATION NUMBER: US/60/234,923
 ; PRIOR FILING DATE: 2000-09-25
 ; PRIOR APPLICATION NUMBER: US/60/235,134
 ; PRIOR FILING DATE: 2000-09-25
 ; PRIOR APPLICATION NUMBER: US/60/235,637
 ; PRIOR FILING DATE: 2000-09-26
 ; PRIOR APPLICATION NUMBER: US/60/235,638
 ; PRIOR FILING DATE: 2000-09-26
 ; PRIOR APPLICATION NUMBER: US/60/235,711
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: US/60/235,720

```

; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 193
; LENGTH: 6465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-193

Alignment Scores:
Pred. No.: 0.0813 Length: 6465
Score: 95.00 Matches: 45
Percent Similarity: 39.43% Conservative: 24
Best Local Similarity: 25.71% Mismatches: 56
Query Match: 9.11% Indels: 50
DB: 10 Gaps: 8

US-10-069-544-2 (1-204) x US-09-954-456-193 (1-6465)
QY 47 MetGlyLeuAspGlnAlaLysGlnGln-----Leu 57
DB 865 ATGATTAACCTATCAAAACATGCGCCAGCAAGCAATTAAGCTGCACTCTATGAGAACTG 924
QY 58 AspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArg----- 72
DB 925 TGTGCACTCTCTATTTCCAAAGATCACTATGATGAGCAATACAAATGTCATCGACGCA 984
QY 73 -----ThreualalysValTyrGlnAlaSerGlnAspAlaThrIleGlnThr 88
DB 985 ATGAAAGAAATTACACGAGCTTACCACTGAAAGTTGCTGAGTGTCTTAAGACAGCT 1044
QY 89 -----LysAlaGlnArgPheGlnLysAla 97
DB 1045 TCTAAGGCTTGTGTAGTAAACGTGAATTTAAGAGCGCAACAGTTAATTAACATGCA 1104
QY 98 IleGluLeu-----AsnProLysAspMetGlnSerTyrMet 109
DB 1105 GTGATATTGTCACCGGATCATTTTGGATCCAAACCCCAAAATATTCGATACACTGCTA 1164
QY 110 AspTyrGlyPheTyrLeuValGlnMetGlyAspLeu-----SerGlyAlaLeuIleTyr 127
DB 1165 GATTAAAGGCTTCACTACTCAATGATTAATATGTCAGTGTGTCCAATTATTCAG 1224
QY 128 PheAspLysProSerArgAlaIleGlyTyrGlnLysArg-----ValAlaAla 143
DB 1225 GCAGCCCTTGACATTGACAGCTAGCTTTGGTGGCAAAATATTCACGTAGCAACAGCT 1284
QY 144 IleGluAsnMetAlaTyr-----IleTyrTyrHisGlnTyrGlnAlaAlaLysSerProThr 162
DB 1285 CATGAAGATTGGCTACTCTCTTATGTCACCACTAATAGCTCGGAAA----- 1335
QY 163 LysAspAspTyrAsnAspAlaLysSerAlaLeuGlnArgAlaLeu 177
DB 1336 -----TTTGACAAATGCACTATTTCATGCAGAAAGACTAATT 1371

RESULT 11
US-10-197-666A-153
; Sequence 153, Application US/10197666A
; Publication No. US20030092037A1
; GENERAL INFORMATION:
; APPLICANT: ASAHI KASEI KABUSIKI KAISYA
; TITLE OF INVENTION: Elki phosphorylation related gene
; FILE REFERENCE: PH-1548US
; CURRENT APPLICATION NUMBER: US/10/197, 666A
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: JP 2001-218204
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: JP 2001-263450
; PRIOR FILING DATE: 2001-08-31
```

```

; PRIOR APPLICATION NUMBER: JP 2002-012176
; PRIOR FILING DATE: 2002-01-21
; PRIOR APPLICATION NUMBER: US 60/305, 884
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/316, 304
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: US 60/350, 027
; PRIOR FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 153
; LENGTH: 3080
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22) ..(2439)
US-10-197-666A-153

Alignment Scores:
Pred. No.: 0.0578 Length: 3080
Score: 92.50 Matches: 35
Percent Similarity: 44.53% Conservative: 26
Best Local Similarity: 25.55% Mismatches: 69
Query Match: 8.87% Indels: 7
DB: 9 Gaps: 4

US-10-069-544-2 (1-204) x US-10-197-666A-153 (1-3080)
QY 28 LysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuAspMet 47
DB 700 AAAAACCAGCCT---ATAGCTATGCTATACAAA-----GGTTAACTTCTTTCACAGA 750
QY 48 GlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPhe 67
DB 751 GGACTTCTGAAGGAAGCAAGCTATTGAATCCTTCAAAGAGCTTTGAAGCAAGAAAGTTGACTT 810
QY 68 AlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAspAlaThrIleGln 87
DB 811 ATTGATGCAATTAAGAAGCTTGGCGCAGCATATAGAACTAGCGCAATTTGAAGCAGCC 870
QY 88 ThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSer 107
DB 871 ACTGAGAC-----TTTCAAAAGCAGCTGTTGCTCAACCAAAATCATGTCGAAACC 921
QY 108 TyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyr 127
DB 922 CTCGAGCTCCGGGGAATGCTCTTACCAACGCGCAGCTTACAGAAAGCCCTTAAGAAC 981
QY 128 PheAspLysProSerArgAlaIleGlyTyrGlnLysArgValAlaIleGlnLysMet 147
DB 982 TTTAAGCGGTGCTGCGACCTAGAGCCATATATGAAGTGTCCAGTATATGAAGGCTC 1041
QY 148 AlaTyrIleTyrTyrHisGln---TyrGlnAlaAlaLysSerProThrLys 163
DB 1042 AGCATGTTGCATGCGACAGATTATGAAGGATTAAGCAAAACAAA 1092

RESULT 12
US-10-197-666A-151
; Sequence 151, Application US/10197666A
; Publication No. US20030092037A1
; GENERAL INFORMATION:
; APPLICANT: ASAHI KASEI KABUSIKI KAISYA
; TITLE OF INVENTION: Elki phosphorylation related gene
; FILE REFERENCE: PH-1548US
; CURRENT APPLICATION NUMBER: US/10/197, 666A
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: JP 2001-218204
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: JP 2001-263450
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: JP 2002-012176
; PRIOR FILING DATE: 2002-01-21
```

```
/ PRIOR APPLICATION NUMBER: US 60/305,884
/ PRIOR FILING DATE: 2001-07-18
/ PRIOR APPLICATION NUMBER: US 60/316,304
/ PRIOR FILING DATE: 2001-09-04
/ PRIOR APPLICATION NUMBER: US 60/350,027
/ PRIOR FILING DATE: 2002-01-23
/ NUMBER OF SEQ ID NOS: 156
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 151
/ LENGTH: 3253
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (9)..(2588)
US-10-197-666A-151

Alignment Scores:
Pred. No.: 0.0628 Length: 3253
Score: 92.50 Matches: 35
Percent Similarity: 44.53% Conservative: 26
Best Local Similarity: 25.55% Mismatches: 69
Query Match: 8.87% Indels: 7
DB: Gaps: 4

US-10-069-544-2 (1-204) x US-10-197-666A-151 (1-3253)
QY 28 LysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMet 47
DB 846 AAAAAACCAAGCCCT--ATAGCTATGCTATACAA-----GGTTTACTTTCTTTCACACA 896
QY 48 GlyLeuAspGlnAlaIleGlnIleuAspAlaIleuSerAlaAspArgGlnPhe 67
DB 897 GSACTCTGAGAGAGACTATGATCCTTCAAGAGAGCTTTGACACAGAAAGTTACTTT 956
QY 68 AlaProAlaTyrArgThrLeuAlaIleValTyrGlnAlaSerGlnAspAlaThrHisGln 87
DB 957 ATTGATGCAATATAAAGCTACGAGGAGCATATAGAGAACTGGCAATTTTGAACACACC 1016
QY 88 ThrLeuAlaGlnArgLeuPheGlnIleuValaIleGlnIleuAsnProLysAspMetGlnSer 107
DB 1017 ACTGAGAGC-----TTTCAAAAGGACACTGTTGCTTCAACCAAAATCATGTGCAAA 1067
QY 108 TyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyr 127
DB 1068 CTCGAGCTCCGGGAGATAGTCTTACCAACGCGACGCTTACAGAGCCCTTAAGAAC 1127
QY 128 PheAspLysProSerArgAlaIleGlyTyrGlnIleValaIleGlnIleuAsnMet 147
DB 1128 TTTAAGCGGTGCTGCAGCTAGACCATATATGAAGTGTGCAATATATGAAGGCGCTC 1187
QY 148 AlaTyrIleTyrTyrHisGln---TyrGlnAlaIleLysSerProThrLys 163
DB 1188 AGCCATGTTGCGATGGAGCATGTTATTAAGGATTAAGCAACAAACAAA 1238

RESULT 13
US-10-108-605-38
/ Sequence 38, Application US/10108605
/ Patent No. US20020160934A1
/ GENERAL INFORMATION:
/ APPLICANT: Broadus, Julie
/ APPLICANT: Stam, Lynn
/ APPLICANT: Bachmann, Jane
/ APPLICANT: Kamdar, Kim
/ TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
/ TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
/ FILE REFERENCE: 3113B
/ CURRENT APPLICATION NUMBER: US/10/108,605
/ CURRENT FILING DATE: 2002-03-27
/ PRIOR APPLICATION NUMBER: US 09/761,142
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/176,418
/ PRIOR FILING DATE: 2000-01-14
/ TYPE: DNA
```

```
/ NUMBER OF SEQ ID NOS: 361
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 38
/ LENGTH: 1994
/ TYPE: DNA
/ ORGANISM: Drosophila melanogaster
US-10-108-605-38

Alignment Scores:
Pred. No.: 0.035 Length: 1994
Score: 92.00 Matches: 29
Percent Similarity: 39.84% Conservative: 20
Best Local Similarity: 23.58% Mismatches: 56
Query Match: 8.82% Indels: 19
DB: Gaps: 2

US-10-069-544-2 (1-204) x US-10-108-605-38 (1-1994)
QY 4 ArgValLysTrpPro----- 8
DB 1042 AGGAGCGAATGGCTTACATCAATCCGAGAGCCGAGAGAGAGAGAGAGGCAACC 1101
QY 9 MetValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProLys 28
DB 1102 TTTTCTTAAAGAAAGGGGAGACTACAGCACCGCTGTAAAGCACTACACCGAGCCATC-AAG 1160
QY 29 AsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGly 48
DB 1161 CGCAATCCGAGATGATCCCAAACTTATATGTAACCGCCCGCTCTACACCAAGCTGCC 1220
QY 49 LysLeuAspGlnAlaIleGlnIleuAspAlaIleuSerAlaAspArgGlnPheAla 68
DB 1221 GCTTTGATCTGGGCTCCCAAGAGTGGACACTGTCATCAAGTACAGAAAGTTACTATA 1280
QY 69 ProAlaTyrArgThrLeuAlaIleValTyrGlnAlaSerGlnAspAlaThrHisGlnThr 88
DB 1281 AAGGCTACATCCCAAGGCGCAAAATTTCTACAGGCGATGCG-----CAACAAATCT 1331
QY 89 LysAlaGlnArgLeuPheGlnIleuValaIleGlnIleuAsnProLysAspMetGlnSerTyr 108
DB 1332 AAGCGCAAGCGCTTACCAAAAGCCCTGAGACTGAGACCCCAACAGCGGAGGCAATT 1391
QY 109 MetAspTyr 111
DB 1392 GAAGGCTAC 1400

RESULT 14
US-10-197-666A-121
/ Sequence 121, Application US/10197666A
/ Publication No. US20030092037A1
/ GENERAL INFORMATION:
/ APPLICANT: ASAHU KASEI KABUSIKI KAISYA
/ TITLE OF INVENTION: E1KI phosphorylation related gene
/ FILE REFERENCE: PH-1548US
/ CURRENT APPLICATION NUMBER: US/10/197,666A
/ CURRENT FILING DATE: 2002-11-18
/ PRIOR APPLICATION NUMBER: JP 2001-218204
/ PRIOR FILING DATE: 2001-07-18
/ PRIOR APPLICATION NUMBER: JP 2001-263450
/ PRIOR FILING DATE: 2001-08-31
/ PRIOR APPLICATION NUMBER: JP 2002-012176
/ PRIOR FILING DATE: 2002-01-21
/ PRIOR APPLICATION NUMBER: US 60/305,884
/ PRIOR FILING DATE: 2001-07-18
/ PRIOR APPLICATION NUMBER: US 60/316,304
/ PRIOR FILING DATE: 2001-09-04
/ PRIOR APPLICATION NUMBER: US 60/350,027
/ PRIOR FILING DATE: 2002-01-23
/ NUMBER OF SEQ ID NOS: 156
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 121
/ LENGTH: 2909
/ TYPE: DNA
```

ORGANISM: Mus musculus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (11) .. (2158)
 US-10-197-666A-121

Alignment Scores:
 Pred. No.: 0.062 Length: 2909
 Score: 92.00 Matches: 48
 Percent Similarity: 41.97% Conservative: 33
 Best Local Similarity: 24.87% Mismatches: 82
 Query Match: 8.82% Indels: 30
 DB: 9 Gaps: 8

US-10-069-544-2 (1-204) x US-10-197-666A-121 (1-2909)

QY 30 AenProGlnLeuAaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
 DB 395 AATGAGAAATTAGCG-----ATTGCTTATGCTCGTGGCGAGCGCTCTC 439
 QY 50 LeuAspGlnAla-----LysGlnGlnLeuAspAla----- 59
 DB 440 TACGATGAAGCCCTAAGACATTTCACAGATGCTCCAGAGGCAATGAGCGCTTCAA 499
 QY 60 ---AlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyr 78
 DB 500 GAAGCTCTGAACAGAAAGTTGACTTATGATGATATATAAGTCTGGAGAGCGCTAC 559
 QY 79 Gln-----AlaSerGlnAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGlyLys 96
 DB 560 AGGAACTGGGCACTTTCAGCA-----GCCACGAGAGCTTCGAGAG 604
 QY 97 AlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuVal 116
 DB 605 GCCCTGCTGCTCAACAGAACCAACGACGAGACCCAGACGCTGGGGATGATGCTCTAC 664
 QY 117 GlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGly 136
 DB 665 CACACGAGAGCCTGCGAGAGGCTTAAAGAACTTCAAGCGCTGCACTGCACTGGAACCG 724
 QY 137 TyrGlnGlyArgValAlaIleGlnAspMetAlaTyrIleTyrThrHisGln---Tyr 155
 DB 725 TACACGAACTGTGCCACGATCAAGAGGGCTGACGATGTGGCCATGGGCGAGTTTAC 784
 QY 156 GlnAlaAlaLysSerProThrLys-----AspAspTyrAsnAsnAlaLysSer 171
 DB 785 GAGGAGTAAAGCAACCAAGTATGCTGATGATGATGATGATGATGATGATGATGATGAT 844
 QY 172 AlaLeuGlnArgAlaLeuIleSerGlyThrGlnHisAspGlu-----IleLysLysSer 189
 DB 845 AGCCAGAGTACCTGAGAGGTCAGTACCTTACGAGAGTATTCGATATTCGACGACAC 904
 QY 190 TyrAspLysLeuLeuSerAspTyrLysLeuLeuSerAsp 202
 DB 905 CTCGATACCCCTCGACGAGATATACATTGACTCGAT 943

RESULT 15
 US-09-917-800A-1526
 ; Sequence 1526, Application US/09917800A
 ; Patent No. US20020119462A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mendrick, Donna
 ; APPLICANT: Porter, Mark
 ; APPLICANT: Johnson, Kory
 ; APPLICANT: Castle, Arthur
 ; APPLICANT: Elashoff, Michael
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Molecular Toxicology Modeling
 ; FILE REFERENCE: 44921-5038-US
 ; CURRENT APPLICATION NUMBER: US/09/917,800A
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: US 60/222,040
 ; PRIOR FILING DATE: 2000-07-31

;; PRIOR APPLICATION NUMBER: US 60/222,880
 ;; PRIOR FILING DATE: 2000-11-02
 ;; PRIOR APPLICATION NUMBER: US 60/290,029
 ;; PRIOR FILING DATE: 2001-05-11
 ;; PRIOR APPLICATION NUMBER: US 60/290,645
 ;; PRIOR FILING DATE: 2001-05-15
 ;; PRIOR APPLICATION NUMBER: US 60/292,336
 ;; PRIOR FILING DATE: 2001-05-22
 ;; PRIOR APPLICATION NUMBER: US 60/295,798
 ;; PRIOR FILING DATE: 2001-06-06
 ;; PRIOR APPLICATION NUMBER: US 60/297,457
 ;; PRIOR FILING DATE: 2001-06-13
 ;; PRIOR APPLICATION NUMBER: US 60/298,884
 ;; PRIOR FILING DATE: 2001-06-19
 ;; PRIOR APPLICATION NUMBER: US 60/303,459
 ;; PRIOR FILING DATE: 2001-07-09
 ;; NUMBER OF SEQ ID NOS: 1740
 ;; SOFTWARE: PatentIn Ver. 2.1
 ;; SEQ ID NO: 1526
 ;; LENGTH: 1632
 ;; TYPE: DNA
 ;; ORGANISM: Rattus norvegicus
 ;; FEATURE:
 ;; OTHER INFORMATION: Genbank Accession No. US20020119462A1 Y15068
 ; US-09-917-800A-1526

Alignment Scores:
 Pred. No.: 0.0484 Length: 1632
 Score: 90.00 Matches: 45
 Percent Similarity: 40.21% Conservative: 33
 Best Local Similarity: 23.20% Mismatches: 80
 Query Match: 8.63% Indels: 36
 DB: 10 Gaps: 7

US-10-069-544-2 (1-204) x US-09-917-800A-1526 (1-1632)

QY 16 ValLeuSerAlaCysGlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGln 35
 DB 463 CTCATAGAGCAATCAACAGAACCAAGCTTCAGACTGGGCAAGAACTCAAGATCCCCGG 522
 QY 36 IleArgThrGlnIleAlaIleSerLeu---LeuAspMetGlyLysLeuAspGlnAlaLys 54
 DB 523 GTCATGACTACTCTCAGATGCTCTGAGTTGATCTGGGCAATGATGATGATGATGATGAT 582
 QY 55 GlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeu 74
 DB 583 GAG-----GCCAGCAACACCCCACTCCACCCCTCTTAAAGAG 624
 QY 75 AlaLysValTyrGlnAlaSerGlnAspAlaThrHisGlnThrLysAlaGlnArgLeuPhe 94
 DB 625 GCCAAGCCAGAACCAATGAGAGAGATCTTCCA---GAGATTAAGAAACAGGCTCTGAA 681
 QY 95 GlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyr 114
 DB 682 GAAAG-----GACCTGGGAAATGAT 702
 QY 115 LeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAla 134
 DB 703 GCTTACAGAGAAAGATTTTGACAGAGCCCTGGAAGCATTTGACAAAG----- 750
 QY 135 IleGlyTyrGlnGlyArgValAlaIleGlnAspMetAlaTyrIleTyrThrHisGln 154
 DB 751 -----GCCAAGAGCTGAGACCTTCAATATGATGATGATGATGATGATGATGATGAT 789
 QY 155 TyrGlnAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlu 174
 DB 790 AATCAAGAGCTGTGCACATTGAGAGAGGCGAATCAACAAATGCCGGAGCTCTGTAG 849
 QY 175 ArgAlaLeuIleSerGlyThrGlnHisAspGlu-----IleLysLysSerTyr 190
 DB 850 AAGGCATTTGAGTGGAGAGAGAACCGAGAGACTACCGTCAGATGCGCCAAAGCTTAT 909
 QY 191 AspLysLeuLeuSerAspTyrLysLeuLeuSerAspTyrLys 204

[illegible][illegible]

```

APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for windows version 4.0
SEQ ID NO 14067
LENGTH: 2230
TYPE: DNA
ORGANISM: Homo sapiens
US-10-198-846-14067

Alignment Scores:
Pred. No.: 0.0776 Length: 2230
Score: 90.00 Matches: 39
Percent Similarity: 38.29% Conservative: 28
Best Local Similarity: 22.29% Mismatches: 78
Query Match: 8.63% Indels: 30
Gaps: 5

US-10-069-544-2 (1-204) x US-10-198-846-14067 (1-2230)

QY 30 AsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
DB 846 AATCCAGACAAATGATATATTAAGTTCTCTCCCTGAGCTTCAAGATGAGACAG 905
QY 50 LeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
DB 906 GAAGCTCAGAGAGAAAAGTACATTGAAGAAGCTCTAGCAACATGCTCCACAGACCTAT 965
QY 70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAspAlaThrHisGlnThrLys 89
DB 966 GTCCTTGATGATGACGCGCAAGTTTACCGAAGAAAGGCTCTGTG-----GATAAA 1016
QY 90 AlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMet 109
DB 1017 GCTCTTGAGTTATTAAGAAAGCCTTCAGAGAAACCCCACTTCTGTCTTACTGCATC 1076
QY 110 AspTyrGlyPhe---TyrLeuValGlnMetGlyAspLeuSerGlyAlaLeu----- 125
DB 1077 CAGATAGGGCTTGTCTCAAGGACCAATGATCCAAATCAAGAGGCTACAAAAGGCGAG 1136
QY 126 -----IleTyrPheAspLysProSerArgAlaIleGlyTyrGlnGly 139
DB 1137 CTTAGAGGCGCAACAGAGAAAGCTAGCAAAATGATGAATAGAGCC----- 1184
QY 140 ArgValValAlaIleGlnAsnMetAlaTyrIleTyrThrHisGlnTyrGlnAlaAlaLys 159
DB 1185 -----ATATTTCATTGATGATGCGAGGAAAA 1214
QY 160 SerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlnAlaLeuIleSer 179
DB 1215 AAGCCACA-----TTGAGGTGGCTCATCTAGACCTGCGCAAGAAATGATATGAA 1265
QY 180 GlyThrGlnHisAspGlnIleLysSerTyrAspLysLeu 194
DB 1266 GCAGGCATCAACAGAAAGCTGAAGAGATTTCAAAATTGTGA 1310

RESULT 19
US-10-193-950A-1
Sequence 1, Application US/10193950A
Publication No. US20030104002A1
GENERAL INFORMATION:
APPLICANT: Sparling, P. Frederick
APPLICANT: Thompson, Stuart
TITLE OF INVENTION: Antigenic from repressible proteins from N. meningitidis related

```

```

TITLE OF INVENTION: hemolysin family of toxins
FILE REFERENCE: 381-39 CIP/CONTI/CPA/CON
CURRENT APPLICATION NUMBER: US/10/193,950A
CURRENT FILING DATE: 2002-12-23
PRIOR APPLICATION NUMBER: 09/045,177
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 3758
TYPE: DNA
ORGANISM: Neisseria Meningitidis
FEATURE:
NAME/KEY: misc feature
LOCATION: (298)..(3642)
OTHER INFORMATION: mat-peptide
FEATURE:
NAME/KEY: CDS
LOCATION: (298)..(3642)
OTHER INFORMATION:
US-10-193-950A-1

Alignment Scores:
Pred. No.: 0.171 Length: 3758
Score: 90.00 Matches: 58
Percent Similarity: 36.40% Conservative: 33
Best Local Similarity: 23.20% Mismatches: 69
Query Match: 8.63% Indels: 90
Gaps: 15

US-10-069-544-2 (1-204) x US-10-193-950A-1 (1-3758)

QY 1 MetLysIleArgValLysTrpProMetValMetAlaMetGlyLeuValLeuSerAlaCys 20
DB 1951 ATGGCACTTCAACCGATTGACGCAACGCGCTAATGAGTATTCACATG----- 2001
QY 21 GlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIle 40
DB 2002 -----ACACCA-----TCCCAAGTAGCACAATAAAGAACGCT 2037
QY 41 AlaIleSerLeuLeuAspMetGlyLys-----LeuAspGlnAlaLysGlnGln----- 56
DB 2038 TTAGTTTCCCTTTCGATTAAGCTTAAGCAGCTATGACGCGCGGACCGCATTCGCC 2097
QY 57 ---LeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAla 75
DB 2098 GTGCTTGATGCTCTACACGGGCGAGATTCACG-----ACACTC--- 2136
QY 76 LysValTyrGlnAlaSerGlnAspAla-----ThrHis 86
DB 2137 ---TATTACATGAGGAGAAAGACCGCTTAATATCGTCAAAAGTAACCAAGATACATAC 2193
QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIle-----GlnLeuAsnProLys 103
DB 2194 GACCATCTCGCAAAACATCTACCAAAACCTGTGTTCACAAACCCGTTTGAGCGCA--- 2250
QY 104 AspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGly----- 119
DB 2251 -----TATTGATCAAAATCAGTTTCAAAATGGA 2280
QY 120 -----AspLeuSerGlyAlaLeuIleTyrPheAsp----- 129
DB 2281 AATGATACGTCTACCTTGTGATTTTGTGTCTTCAAGCATTTAACCATCTCAAGAA 2340
QY 130 ---LysProSerArgAlaIle----- 135
DB 2341 ACTAATCCGCAAAAAGCTTTTGTGATTTGGCCGAGATGCTTCATATGAGCAACTTCGT 2400
QY 136 ---GlyTyrGlnGlyArgValValAlaIleGlnAsnMetAlaTyrIleTyrThrHisGln 154
DB 2401 TCTGTGTAGGAAGCGGAGAGACTAATGCGCAT-----TATGTG 2439
QY 155 TyrGlnAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGln 174

```



```

Qy      88 ThrlyAlaGlnArgLeuPheGlnIuylsAlaIleGluLeuAsnProLysAspMetGlnSer 107
      359 GATATTGGCCAGGGCCATTGTAAGGTCACAGAAAGCATATCCCGATGATCTTGAAGCT 418
Qy      108 TyrMetAspTyrGlyPheTyrIleuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyr 127
      419 TGGATTGAAATGGCA--CAAACTTAGAACAGACTGATATACAGGGTCCCTTCACGCC 475
Db      128 PheAspLysProSerArgAlaIleGlyTyrGlnGlyArgValValAla----- 143
      476 TATGACACAGCAACGAACTCTT-----CAGGAAAGATGACAGCGCATCTCTCCA 529
Qy      144 -----IleGluAsnMetAlaTyrIleTyrTyrHis----- 153
      530 GAGATTCTCAATTAATGTGGTGCCCTCCATTGACCTTGAACCTAGAGGAGGCTAAG 589
Qy      154 -----GlnTyrGlnAlaAlaLysSerProThrLysAspAsp-----Tyr 166
      590 AAATATTTTGGCGCTCATTTGACCGCTGCMAAAGCAGAACCGGACAGATGAGCATTAAC 649
Db      167 AsnAsnAlaLysSerAla-----LeuGluArgAlaLeuIleSerGlyThr 181
      650 TATAACGCATTTCCGTTACCAAGCTCATATATCTCCGACGCTTATGAGCGCATGTGT 709
Qy      182 GlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSerAspTyr 197
      710 GAATTCATGAAGCAGAAAACCTGATATAAAACATCTTACGCGAATCAT 757

```

RESULT 24
US-10-114-170-242/c
Sequence 242, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. US20030023075A1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Charles & Brady
Street: 1 South Plinkney Street
City: Madison
State: WI
Country: US

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 242:
SEQUENCE CHARACTERISTICS:
LENGTH: 31880
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 242:
US-10-114-170-242

Alignment Scores:
Pred. No.: 11.2 Length: 31880
Score: 87.00 Matches: 49
Percent Similarity: 38.03% Conservative: 40
Best Local Similarity: 20.94% Mismatches: 77
Query Match: 8.34% Indels: 68
Gaps: 10

US-10-069-544-2 (1-204) x US-10-114-170-242 (1-31880)

```

Qy      21 GlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIle 40
      3088 AAAGCTACAAACCGATTATCTGGAATACAGAAACAAATGCTGGAAGTGGT----- 3038
Qy      41 AlaIleSerLeuAspMetGly-----LysLeuAspGlnAlaLys 54
      3037 -----GAAGTCTGGAATCCGAGCTGCAAAATGTTAGTGCATCTCTTGACCAACACGC 2984
Db      55 GlnGlnLeuAspAlaAlaLeuSerAlaAsp----- 64
      2983 CAGCAAAATA-----CTTCTTAATGACAGATTCAATGCGCGTGGCCCTGATGAAGTACC 2930
Qy      65 -----ArgGlnPheAlaProAlaTyrArgThrLeuAlaLysVal 77
      2929 GAAGATTACAGACATCGCTTACCCAAACACACAGACATTTGGCTCGATGGACATGAT 2870
Db      78 TyrGlnAspSerGluAspAlaThrHisGlnThrLysAlaGlnArgLeu----- 93
      2869 GTACAGCAATTCCTCCGAAACCTTAACCAAGCAGCTTAGGTGAACATTAATCGAAATGAA 2810
Qy      94 -----PheGluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyr 108
      2809 CAAAGTGCGAAGAAATTCCTGAATACCTAGACGCTCGAAAGAAATTCGATAGAAAT 2750
Db      109 MetAspTyrGly---PheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle--- 126
      2749 ATGGAACAAATTAACAGCAATGATGATGTTGTCAGTAAGACTGTGAAAGATTAAC 2690
Qy      127 -----TyrPheAspLys 130
      2689 CATCACTATCCAGTAATCGCATATATGTTGATTCATCCAGACATCATTTGATAG 2630
Qy      131 ProSerArgAlaIleGlyTyrGlnGlyArgValAlaIleGluAsnMetAlaTyrIle 150
      2629 GCAAGTGCAGGCGCTGACTTCGACAGTCAGA---GAATCAATGAAAATTTGCTCTATCC 2573
Qy      151 TyrTyrHisGlnTyrGluAlaAlaLysSerProThrLysAspAspTyrAsnAsnAlaLys 170
      2572 ATCAACGACAGCTTACATGCTTTTGACCAACCACTGAACCTGAATGCAACCTGGAATG 2513
Db      171 SerAlaLeuGluArgAlaLeuIleSerGlyThrGlnHisAspGluIleLysLys----- 188
      2512 CAATCATTAAGGTAATGCTCTGCTTCA-----ATCAGCAAAAGCTTTT 2471
Qy      189 -----SerTyrAspLysLeuLeuSerAspTyrLysLeuLeu 200
      2470 GTCGTAACATAAGAAAACCTTATTAAGATTACCAAAATAGTT 2429

```

RESULT 25
US-10-128-714-2166
Sequence 2166, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wengqi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M

```

APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: Patent in version 3.1
SEQ ID NO 2166
LENGTH: 2031
TYPE: DNA
ORGANISM: Aspergillus fumigatus
US-10-128-714-2166

Alignment Scores:
Pred. No.: 0.236 Length: 2031
Score: 86.00 Matches: 44
Percent Similarity: 41.99% Conservative: 32
Best Local Similarity: 24.31% Mismatches: 73
Query Match: 8.25% Indels: 32
DB: 9 Gaps: 8

US-10-069-544-2 (1-204) x US-10-128-714-2166 (1-2031)
QY 28 LysAsnAspProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMet 47
DB 1135 AAGGATGTGATGCGGCGCGGAGACGCTGGCCCAAGCCATTGGATGCGCCCAAGACAG 1191
QY 48 GlyLysLeuAspGlnAlaIleGlnLeuAspAlaIleLeuSerAlaAspArgGln--- 66
DB 1192 AAGAAATTCACATTTGCGCAAGATCTGGCTGTAAAGCGCAATTCGACATCGCGCAGATG 1251
QY 67 ---PheAlaProAlaTyrArgThrLeuAla-----Lys 76
DB 1252 GATCTGCAGCGCTGCGGGAAGACGCTGGCCCAAGCCATTGGATGCGCCCAAGACAG 1311
QY 77 ValTyrGlnAlaSerGluAspAlaThrHisGln-----ThrLysAlaGlnArg 92
DB 1312 CTCTCCGGGGTTATATGATCTCGAGCGGAGCTGTTGAGTGTGCGGCGGAGC 1371
QY 93 LeuPheGluLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112
DB 1372 CTGTACGAAAAACAGATCGAATGGAACCGGCGCAATCGCATCGAATCAAGTACGCT 1431
QY 113 PheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSer 132
DB 1432 GAGCTGCAGCGAGACTGATGATCTCGGAACGTCGCGCGGCGCATCTTCCAG----- 1482
QY 133 ArgAlaIleGlyTyrGlnGluValValAlaIleGluAsnMet-----AlaTyr 149
DB 1483 -----CTCGGATCATCGACCGATGTCGATATCCGAGCTGCTCGAAGCGTAC 1536
QY 150 IleTyrTyrHisGlnTyrGlnAlaAlaLysSerProThrLysAspLysPheAsnAla 169
DB 1537 ATTGACTTTGAAGAGTACAAAGC-----GAGTACACACGAGTGC 1575
QY 170 LysSerAlaLeuGlnArgAlaLeuIleSerGlyThrGlnHisAspGlnIleLysLysSer 189
DB 1576 CCGCAGCTTACGAGCGT---CTTTGCGAAGACAGACATGTCAGAGTGTGATCAAC 1632
QY 190 Tyr 190
DB 1633 TAC 1635

```

```

RESULT 26
US-10-128-714-7166
Sequence 7166, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wengdi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: Patent in version 3.1
SEQ ID NO 7166
LENGTH: 2031
TYPE: DNA
ORGANISM: Aspergillus fumigatus
US-10-128-714-7166

Alignment Scores:
Pred. No.: 0.236 Length: 2031
Score: 86.00 Matches: 44
Percent Similarity: 41.99% Conservative: 32
Best Local Similarity: 24.31% Mismatches: 73
Query Match: 8.25% Indels: 32
DB: 9 Gaps: 8

US-10-069-544-2 (1-204) x US-10-128-714-7166 (1-2031)
QY 28 LysAsnAspProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMet 47
DB 1135 AAGGATGTGATGCGGCGCGGAGACGCTGGCCCAAGCCATTGGATGCGCCCAAGACAG 1191
QY 48 GlyLysLeuAspGlnAlaIleGlnLeuAspAlaIleLeuSerAlaAspArgGln--- 66
DB 1192 AAGAAATTCACATTTGCGCAAGATCTGGCTGTAAAGCGCAATTCGACATCGCGCAGATG 1251
QY 67 ---PheAlaProAlaTyrArgThrLeuAla-----Lys 76
DB 1252 GATCTGCAGCGCTGCGGGAAGACGCTGGCCCAAGCCATTGGATGCGCCCAAGACAG 1311
QY 77 ValTyrGlnAlaSerGluAspAlaThrHisGln-----ThrLysAlaGlnArg 92
DB 1312 CTCTCCGGGGTTATATGATCTCGAGCGGAGCTGTTGAGTGTGCGGCGGAGC 1371
QY 93 LeuPheGluLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112
DB 1372 CTGTACGAAAAACAGATCGAATGGAACCGGCGCAATCGCATCGAATCAAGTACGCT 1431
QY 113 PheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSer 132
DB 1432 GAGCTGCAGCGAGACTGATGATCTCGGAACGTCGCGCGGCGCATCTTCCAG----- 1482
QY 133 ArgAlaIleGlyTyrGlnGluValValAlaIleGluAsnMet-----AlaTyr 149
DB 1483 -----CTCGGATCATCGACCGATGTCGATATCCGAGCTGCTCGAAGCGTAC 1536

```



```

Db      1192 AAGAAATTCACATTGCCAAGATCTGGCTTTAAAGCGCAATTGCACATCCGCCAGATG 1251
Qy      67 ---PhealAproalAlyrgrThrlleuAa-----Lys 76
Db      1252 GATTCGACAGGCTGGCGGAAAGACGCTGGGCAAGCCATTGGATGTGCCCAAGAGCAAG 1311
Qy      77 ValTYrGlnAlaSerGlnAspAlaThrHisGln-----ThrlsAlaGlnArg 92
Db      1312 CTTTCGGGGGTTATATTGATCTCGAAGCGGAGCTGTTTGAAGTGTGGCGGTGGCGGACC 1371
Qy      93 LeupheGluLYsAlaIleGluLeuAsnProLYsAspMetGlnSerTYrMetAspTYrGly 112
Db      1372 CTGTACGAAAAACAGATCGAATGGAACCCGGCCCATCGCATCGATCAAGTACGCT 1431
Qy      113 PheTYrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTYrPheAspLYsProSer 132
Db      1432 GAGCTGAGACGAGACCTGATGATCTCGAAGCGTGGCGGCGATCTTCAG----- 1482
Qy      133 ArgAlaIleGlyTYrGlnGlyArgValAlaIleGluAsnMet-----AlaTYr 149
Db      1483 -----CTCGGATCGATAGCCGATGCTGATATGCGGACCTGCTCGAAGGGGTC 1536
Qy      150 IleTYrTYrHisGlnTYrGlnAlaAlaLYsSerProThrlsAspAspTYrAsnAsnAla 169
Db      1537 ATTGACTTTGAAGAGTACGAGGCG-----GAGTACGACCGAGTGTG 1575
Qy      170 LysSerAlaLeuGlnArgAlaLeuIleSerGlyThrlHisAspGlnIleLYsLYsSer 189
Db      1576 CGGACGCTGTACGAGCGT---CTTTTCGAAAGACAGACCATGTCAAGTGTGATCAAC 1632
Qy      190 TYr 190
Db      1633 TAC 1635

RESULT 29
US-10-128-714-166
; Sequence 166, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 166
; LENGTH: 4082
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-166

Alignment Scores:
Pred. No.: 0 68          Length: 4082
Score: 86.00          Matches: 44
Percent Similarity: 41.99%          Mismatches: 32
Best Local Similarity: 24.31%          Matches: 73

```

```

Query Match: 8.25%          Indels: 32
DB: 9          Gaps: 8

US-10-069-544-2 (1-204) x US-10-128-714-166 (1-4082)
Qy      28 LysAsnAsnProGlnLeuAlaGlnIleArgThrlGlnIleAlaIleSerLeuAspMet 47
Db      2135 AAGATGTGATTCGGGGCCCGGCAGATCTACACGGA---TGCCTGAAGCTATCCACAT 2191
Qy      48 GlyLYsLeuAspGlnAlaLYsGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGln--- 66
Db      2192 AAGAAATTCACATTGGCCAAATCTGCTTTAAAGCGCAATTGCACATCCGCAAGTGTG 2251
Qy      67 ---PhealAproalAlyrgrThrlleuAa-----Lys 76
Db      2252 GATTCGACAGGCTGGCGGAAAGACCTGGGCCCAAGCCATTGGGATGTGCCCAAGAGCAAG 2311
Qy      77 ValTYrGlnAlaSerGlnAspAlaThrHisGln-----ThrlsAlaGlnArg 92
Db      2312 CTTTCGGGGGTTATATTGATCTCGAAGCGGAGCTTTGAGTTCGTGCGGTGGCGGACC 2371
Qy      93 LeupheGluLYsAlaIleGluLeuAsnProLYsAspMetGlnSerTYrMetAspTYrGly 112
Db      2372 CTGTACGAAAAACAGATCGAATGGAACCCGGCCCATGCGCATGTGATCAAGTACGCT 2431
Qy      113 PheTYrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTYrPheAspLYsProSer 132
Db      2432 GAGCTGAGACGAGACCTGATGATCTCGAAGCTGCCGGGAGATCTTCAG----- 2482
Qy      133 ArgAlaIleGlyTYrGlnGlyArgValAlaIleGluAsnMet-----AlaTYr 149
Db      2483 -----CTCGGATCGATACCCGATGCTGATATGCGGACCTGCTCGAAGGGGTC 2536
Qy      150 IleTYrTYrHisGlnTYrGlnAlaAlaLYsSerProThrlsAspAspTYrAsnAsnAla 169
Db      2537 ATTGACTTTGAAGAGTACGAGGCG-----GAGTACGACCGAGTGTG 2575
Qy      170 LysSerAlaLeuGlnArgAlaLeuIleSerGlyThrlHisAspGlnIleLYsLYsSer 189
Db      2576 CGGACGCTGTACGAGCGT---CTTTTCGAAAGACAGACCATGTCAAGTGTGATCAAC 2632
Qy      190 TYr 190
Db      2633 TAC 2635

RESULT 30
US-10-128-714-5166
; Sequence 5166, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 5166
; LENGTH: 4082
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5166

Alignment Scores:
Pred. No.: 0.68 Length: 4082
Score: 86.00 Matches: 44
Percent Similarity: 41.99% Conservative: 32
Best Local Similarity: 24.31% Mismatches: 73
Query Match: 8.23% Indels: 32
DB: 9 Gaps: 8

US-10-069-544-2 (1-204) x US-10-128-714-5166 (1-4082)

QY 28 LysAsnAnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMet 47
DB 2135 AAGGATGTGATCGGCGCCCGCCAGATCTACACGGAA---TGCCTGAAGCTGATCCACAT 2191
QY 48 GlyLeuLeuAspGlnAlaIleGlnIleuAspAlaAlaLeuSerAlaAspArgIn--- 66
DB 2192 AAGAAATTCACATTTGCCAAGATCTGCTTTAAAGCGCAATTGCACATCCGACATG 2251
QY 67 ---PheAlaProAlaIleArgThrLeuAla-----Lys 76
DB 2252 GATCGCAGAGCTGCGCGAAGCGCTGGGCCAAGCATTTGGATGCCCCAAGACAAAG 2311
QY 77 ValTyrGlnAlaSerGlnuAspAlaThrIleGln-----ThrLysAlaGlnArg 92
DB 2312 CTCTCCGGGGTTATATGATCTCGACGGCGAGCTGTTGAGTTGCTGCGGCGCGGACC 2371
QY 93 LeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112
DB 2372 CTGTACGAAACACATGCAATGCAACCGGCCCAATAGCCATCTGATCAAGTACGCT 2431
QY 113 PheTyrLeuValGlnMetGlyAspLeuSerGlnAlaLeuIleTyrPheAspLysProSer 132
DB 2432 GAGCTGAGCAGCAGCATGATGATCTCGAAGCGTGGCGGCGCATCTTCGAG----- 2482
QY 133 ArgAlaIleGlyTyrGlnGlyArgValAlaIleGlnuAsnMet-----AlaTyr 149
DB 2483 -----CTCGGGATGATCAGCCGATGCTGATATGCGGAGCTGCTCGAAGCGTAC 2536
QY 150 IleTyrTyrHISGlnTyrGlnAlaAlaLysSerProThrLysAspAspTyrAsnAla 169
DB 2537 ATTGCTTTGAAGATGACGAAGC-----GAGTACGACCGAGTGT 2575
QY 170 LysSerAlaLeuGlnArgAlaLeuIleSerGlyThrGlnHISAspGlnIleLysLysSer 189
DB 2576 CGGACGCTGTACGACGCT---CTTTTGCAAGAACACAGACATGTCAAGTGTGATCAAC 2632
QY 190 Tyr 190
DB 2633 TAC 2635

RESULT 31
US-10-198-846-9055
; Sequence 9055, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; FILE REFERENCE: THERAPY OF BREAST CANCER
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18

```

```

; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9055
; LENGTH: 496
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2, 3, 4, 5, 6, 7, 8, 9, 35, 40, 42, 43, 45, 48, 63, 67, 72,
; LOCATION: 73
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9055

Alignment Scores:
Pred. No.: 0.0326 Length: 496
Score: 85.50 Matches: 22
Percent Similarity: 52.44% Conservative: 21
Best Local Similarity: 26.83% Mismatches: 35
Query Match: 8.20% Indels: 4
DB: 9 Gaps: 1

US-10-069-544-2 (1-204) x US-10-198-846-9055 (1-496)

QY 31 ProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeu 50
DB 245 CCTTCTCAGAGGAGAGCTCCGCTTGACATGCTCAGGTTTGCCGCTGATGGTCAACA 304
QY 51 ArgGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAla 70
DB 305 AAGAGCTGAAGATGACCAATCATTTGTTCAGAGGACCGGATGCTTGAAATGC 364
QY 71 TyrArgThrLeuAlaLysValTyrGlnAlaSerGlnuAspAlaThrIleGlnThrLysAla 90
DB 365 TATGCCCTCTGTGACCATCTATACGACAGAG-----AACGACAGACAGCA 415
QY 91 GlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAsp 110
DB 416 CTGATGCTATACACAGAGCTCTCAGCTGAACCAAGACCCAAAGTCAATTTCTGAA 475
QY 110 P Tyr 111
DB 476 CTTT 479

RESULT 32
US-09-938-842A-2043
; Sequence 2043, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2043
; LENGTH: 1128
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2043

Alignment Scores:
Pred. No.: 0.113 Length: 1128
Score: 85.50 Matches: 47

```

Percent Similarity:	38.5%	Conservative:	34
Best Local Similarity:	22.3%	Mismatches:	70
Query Match:	8.2%	Indels:	59
DB:	9	Gaps:	10

US-10-069-544-2 (1-204) X US-09-938-842A-2043 (1-1128)

[illegible]

```

RESULT 33
US-09-880-107-2411
/ Sequence 2411, Application US/09880107
/ Patent No. US20020142981A1
/ GENERAL INFORMATION:
/ APPLICANT: Horne, Darci T.
/ APPLICANT: Vockley, Joseph G.
/ APPLICANT: Scherf, Iwe
/ APPLICANT: Gene Logic, Inc.
/ TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
/ FILE REFERENCE: 44921-5028-WO
/ CURRENT APPLICATION NUMBER: US/09/880,107.
/ PRIOR FILING DATE: 2001-06-14
/ PRIOR APPLICATION NUMBER: US 60/211,379
/ PRIOR FILING DATE: 2000-06-14
/ PRIOR APPLICATION NUMBER: US 60/237,054
/ PRIOR FILING DATE: 2000-10-02
/ NUMBER OF SEQ ID NOS: 3950
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2411
/ LENGTH: 2113
/ TYPE: DNA
/

```

; ORGANISM: Homo sapiens.
 ;
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981a1 M86752
 ;
 US-09-880-107-2411

Alignment Scores:

Pred. No.:	0.401	Length:	2113
Score:	84.50	Matches:	44
Percent Similarity:	38.14%	Conservative:	30
Best Local Similarity:	22.68%	Mismatches:	67
Query Match:	8.10%	Indels:	53
DB:	10	Gaps:	9

US-10-069-544-2 (1-204) X US-09-880-107-2411 (1-2113)

[illegible]

RESULT 34
 US-09-938-842A-5
 : Sequence 5, Application US/09938842A
 : Patent No. US20020160378A1
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Harper, Jeff
 : APPLICANT: Kreps, Joel
 : APPLICANT: Wang, Xun
 : APPLICANT: Zhu, Tong
 :
 : TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 : FILE REFERENCE: SAME, AND METHODS OF USE
 : FILE REFERENCE: SCRIPI300-3
 : CURRENT APPLICATION NUMBER: US/09/938,842A
 : CURRENT FILING DATE: 2001-08-24
 : PRIOR APPLICATION NUMBER: US 60/227,866
 : PRIOR FILING DATE: 2000-08-24
 : PRIOR APPLICATION NUMBER: US 60/264,647
 : PRIOR FILING DATE: 2001-01-16
 :

```

; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5
; LENGTH: 4677
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-5

Alignment Scores:
Pred. No.: 1 34 Length: 4677
Score: 84.50 Matches: 45
Percent Similarity: 35.56% Conservative: 19
Best Local Similarity: 25.00% Mismatches: 63
Query Match: 8.10% Indels: 53
DB: 9 Gaps: 7

US-10-069-544-2 (1-204) x US-09-938-842A-5 (1-4677)

Qy 45 LeuAapMetGlyLeuAapGlnAlaLysGlnLeuAapAlaLeu----- 61
Db 2425 CTTGACAAAGGAACTAGATGCTGAGCTATGAGCAAGCAAGCATTTAGTCAAGT 2484
Qy 62 -----SerAlaAapArgGlnPheAlaProAlaTyrArgThrLeuAlaLys 76
Db 2485 ATTGCTGCTGTGGGCTTATCATAGAAATAGTCATGCTAGCAGTCTTCCAGTT 2544
Qy 77 ValTyrGlnAlaSerGlnAapAlaThrHisGlnThrLysAlaGlnArgLeuPheGlnLys 96
Db 2545 GACTATATCCACACCTGGCGATTTTAACGACCAACATATATATCA-----CAGAA 2595
Qy 97 AlaIleGlnLeuAap-----ProLysAapMetGlnSerTyr 108
Db 2596 GCTTAAATATATTAATAGAGAGAACTGTGTCGACATCCGACATATGAAAGCAT 2655
Qy 109 MetAapTyrGlyPheTyrLeuValGlnMetGlyAapLeuSerGlyAlaLeuLysTyrPhe 128
Db 2656 GGGGATCTTCTGTGATTTACTATGCTCTCAACATTTTGAAATGGCAATAAATATG 2715
Qy 129 AapLys-----ProSerArgAlaIle 135
Db 2716 AACCGGGGCTGTTCTTCTTCACTTCACTTGGGCTATCTCACCAAAATCTGCTCC 2775
Qy 136 GlyTyrGlnGly-----ArgValAlaAlaIleGlnAapMetAlaTyrIle 150
Db 2776 ACATATATCATATGCTGCTATGATGAGAAAGAGAGATGAGATCATCTAGCTTAAG 2835
Qy 151 TyrTyrHisGlnTyrGlnAlaAlaLysSerProThrLysAapAapTyrAsnAsnAlaLys 170
Db 2836 TACCTGCAT-----GAAGCTCTAAAGAGCAATAAAGA----- 2868
Qy 171 SerAlaLeuGlnLysAlaLeuLysSerGlyThrGlnHisAapGlnLysLysSerTyr 190
Db 2869 -----TTGCTAGAGCTGACATATTCAGACTGGCGCAAGCTAT 2907

RESULT 35
US-09-978-244A-7
; Sequence 7, Application US/09978244A
; Publication No. US20030103992A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter S
; APPLICANT: Garman, Jonathan D.
; APPLICANT: Candia III, Albert F.
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP MEMBRANE PROTEINS
; FILE REFERENCE: 020554-000161US
; CURRENT APPLICATION NUMBER: US/09/978,244A
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/310,028
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 09/737,246
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,969

```

```

; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,960
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,968
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/240,545
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 6225
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5997)
; FEATURE:
; OTHER INFORMATION: CLASP-3
; US-09-978-244A-7

Alignment Scores:
Pred. No.: 2 06 Length: 6225
Score: 84.50 Matches: 44
Percent Similarity: 35.75% Conservative: 25
Best Local Similarity: 22.80% Mismatches: 71
Query Match: 8.10% Indels: 53
DB: 9 Gaps: 10

US-10-069-544-2 (1-204) x US-09-978-244A-7 (1-6225)

Qy 47 MetGlyLysLeuAapGlnAlaLysGlnLeuAapAlaLeuSerAlaAapArgGln 66
Db 4906 GTTGACCAATTCAGAAAGCGGCTGCTCTTCCATGAGG---GGCATGACAAAGCA 4962
Qy 67 PheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAapAla----- 84
Db 4963 GTTATATGAGCTTTCAAAGTACTCTATTCCTATTCATGATGAGTATCCGAGTCAAAGAG 5022
Qy 85 -----ThrHisGlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAapPro 102
Db 5023 CTCTCCACAATTCATGAGCAAACTTCAGAGGCAATTCAGCAAAATCTCATCAGAGATGCT 5082
Qy 103 LysAapMet---GlnSerTyrMetAapTyrGlyPheTyrLeuValGlnMetGlyAapLeu 121
Db 5083 AAGGGAGATTTGGACCTATTTCCGTGTTGTTTATGGAACCAAGTTGGGGGATTTG 5142
Qy 122 SerGlyAlaLeuLysTyrPheAapLysProSer----- 132
Db 5143 GATGAACAGAAATTTGTTTACAGAGACCTTCATTAACCAATCCGACAGATTTCTCAC 5202
Qy 133 ArgAlaIleGlyTyrGlnGlyArg-----ValValAlaIle----- 144
Db 5203 AGATTGAGGAGATTTATGAGAAAGATTGAGAGAGATGCTGAAGTATCATCAGAGAC 5262
Qy 145 -----GlnSerMetAlaTyrIle----- 150
Db 5263 TCTATCTGTAGCAAGTGCAATTAAGTCCAAAGGAGATATTCAGATTACCTAC 5322
Qy 151 -----TyrTyrHisGlnTyrGlnAlaAlaLysSerProThrLysAapAap----- 165
Db 5323 GTGAGGCCCTTCTTGAACCTATGAGATGAGAGACAGAAATCATTTTGAACAAAAT 5382
Qy 166 TyrAsnAsnAlaLysSerAlaLeuGlnArgAlaLeu-----IleSer 179

```



```

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 9510
LENGTH: 3540
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(3540)
US-09-815-242-9510

```

```

Alignment Scores:
Pred. No.: 1.02 Length: 3540
Score: 84.00 Matches: 35
Percent Similarity: 41.36% Conservative: 32
Best Local Similarity: 21.60% Mismatches: 73
Query Match: 8.05% Indels: 22
DB: 10 Gaps: 4

```

US-10-069-544-2 (1-204) x US-09-815-242-9510 (1-3540)

```

OY 22 SerThrProIleProProlYAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAla 41
DB 1225 AGTCGTGACGTTTTCACAAAACAGACAGATCACTAGAAAGCTGAAGCAATTGGCT 1284
OY 42 ILeSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnGlnLeuAspAlaLeu 61
DB 1285 ATAGCT-----AAGAGAGAGCTAGTCAAGAAAGAGAGCTGAAAGCTGCAAG 1335
OY 62 SerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSer 81
DB 1336 GAGCAGCTTCAGAAATTATTGGCTCACTACCAAGCTATTGCCAG--GAGCAAGAGAG 1392
OY 82 GluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGluLysAlaIleGluLeuAsn 101
DB 1393 CAGAAACTCTTCAAGCTCAACAAAGCAAGCAAGCTTGAACCGCTGATATCTCAA 1452
OY 102 ProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeu 121
DB 1453 AACAAGCAGCGCCAGAGCTCAAGATTGGAATAATCTGAGAAATCATAGTAATTTAT 1512
OY 122 SerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArgVal 141
DB 1513 GCAGCTTGAAGAGCTTCTCCAGAAAAAGATCCCTAGGCGGATTAATTTGGTCAATC 1572
OY 142 ValAlaIleGluAsnMetAlaTyrIleTyrHisGlnTyrGluAlaAlaLysSerPro 161
DB 1573 -----ACTGAGCATCTGACCTTGATATGATATCA----- 1605
OY 162 ThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluAlaLeuIleSerGlyThr 181
DB 1606 -----ACTGCTTAAGAAATTTGCACTTGAGCGAGTAGT 1638
OY 182 GlnHis 183
DB 1639 CAGCAC 1644

```

RESULT 39
US-10-106-698-2038

```

Sequence 2038, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 2038
LENGTH: 1783
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (27)..(27)
NAME/KEY: misc_feature
LOCATION: (94)..(95)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (97)..(97)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1576)..(1576)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1692)..(1692)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1694)..(1694)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1733)..(1733)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1747)..(1747)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (1772)..(1772)
OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-2038

```

```

Alignment Scores:
Pred. No.: 0.424 Length: 1783
Score: 83.50 Matches: 32
Percent Similarity: 38.92% Conservative: 33
Best Local Similarity: 19.16% Mismatches: 63
Query Match: 9.01% Indels: 39
DB: 9 Gaps: 4

```

US-10-069-544-2 (1-204) x US-10-106-698-2038 (1-1783)

```

OY 35 GlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLys 54
DB 844 GAGTACGCCATTAACCTAGAGATTGTGCTACATATCACTGAAGCAAGCTTCAACAGCACAA 903
OY 55 GlnIleuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeu 74
DB 904 GACCACTTGCAACATGCCCTGAATCTTAATAGGACAGATCG--ACTTATATATGCTG 960
OY 75 AlaLysValTyrGlnAlaSerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPhe 94
DB 961 GGGAGATCACTTGCTGAGGAGGACTTG-----GACAAAGGCATTGAAGCTTAC 1011
OY 95 GluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyr 114

```

```

Db      1012 AAGAAAGAGTGGATCTTCAACGAAATATACAGAGCTTTTACCACTTAGGATTC 1071
Qy      115 LeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyPheAspLysProSerArgla 134
Db      1072 TACTTACGCTCGGC----- 1086
Qy      135 IleGlyTyrgLuglyArgValAlaIleGlnMetAlaTyrlleTyTyrlHisGln 154
Db      1087 -----ATTACAGAGGCA 1101
Qy      155 TyrGluAlaAlaLysSerProThryLysAspGlyTyrAsnAlaLysSerAlaLeu 174
Db      1102 TTTGAAACATCTTGGCAATGCACCTGACTTATGACCTTACCACTACAGGCCATCTTGCA 1161
Qy      175 ArgAlaLeuIleSerGlyThrgLinhIsAspGluIleLysSerTyTyrAspLysLeu 194
Db      1162 GCAGCGACGATGATGACAGACCCAGCGGAC-----TTTGATCTTGCCTC 1206
Qy      195 SerAspTyTyrLysLeuLeuSer 201
Db      1207 ACCAATATACAGAGTGTGCT 1227

```

```

RESULT 40
US-10-153-668-193
/ Sequence 193, Application US/10153668
/ Publication No. US20030092616A1
/ GENERAL INFORMATION:
/ APPLICANT: HONDA, Goichi
/ APPLICANT: MATSUDA, Akio
/ APPLICANT: MURAMATSU, Shuji
/ APPLICANT: ISHIZAMA, Kenya
/ TITLE OF INVENTION: STAT6 Activating Gene
/ FILE REFERENCE: 1254-0207P
/ CURRENT APPLICATION NUMBER: US/10/153,668
/ CURRENT FILING DATE: 2002-05-24
/ PRIOR APPLICATION NUMBER: US 60/293,172
/ PRIOR FILING DATE: 2001-05-25
/ PRIOR APPLICATION NUMBER: US 60/316,031
/ PRIOR FILING DATE: 2001-08-31
/ PRIOR APPLICATION NUMBER: US 60/328,403
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: JP 2001-157043
/ PRIOR FILING DATE: 2001-05-25
/ PRIOR APPLICATION NUMBER: JP 2001-260681
/ PRIOR FILING DATE: 2001-08-30
/ PRIOR APPLICATION NUMBER: JP 2001-313175
/ PRIOR FILING DATE: 2001-10-10
/ NUMBER OF SEQ ID NOS: 488
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 193
/ LENGTH: 1731
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (292) ..(1443)
US-10-153-668-193

```

```

Alignment Scores:
Pred. No.: 0.474      length: 1731
Score: 83.00      Matches: 42
Percent Similarity: 38.73%      Conservative: 25
Best Local Similarity: 24.28%      Mismatches: 90
Query Match: 7.96%      Indels: 16
DB: 9      Gaps: 6

```

US-10-069-544-2 (1-204) x US-10-153-668-193 (1-1731)

```

Qy      31 ProGlnLeuAlaGlnIleArgThrgLinhIsSerLeuLeuAspMetGlyLysLeu 50
Db      913 CCGAGAGTCTCTGCGACAGCAAGAAATGTTTGCATTGTACCGCCAGGCATATACG 972

```

```

Qy      51 AspGlnAlaLysGlnGlnLeuAspAlaIleLeuSerAlaAspArgGlnPheAlaProAla 70
Db      973 GGAACCACTCTTCAACAAATCCAGCGATATGAAAGCTTTTGAAGAGGCTATTAAGAAA 1032
Qy      71 TyrArgThrLeuAlaLysValTyr-----GlnAlaSerGluAspAlaThrHis 86
Db      1033 TTTCCAAAGTGTGCCAAGGCTATGACCTATACGCCAGGCCATTAAACAGATCAACAACAG 1092
Qy      87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGln 106
Db      1093 TTTGGTAAAGCTGATGAAATGATGATTAATGATTAATGATTAATGATTAATGATTAATG 1152
Qy      107 SerTyTyrMetAspTyrgLysPheTyrlleValGln---MetGlyAspLeuSerGlyAlaLeu 125
Db      1153 ACATATGTTCATTAAGGTTTACTTCACTTCACTGAGTGAAGCAAGATCTCGATAGGTTTG 1212
Qy      126 IleTyPheAspLysProSerArgAlaIleGlyTyrgLinhIsAspGluIleArgVal---ValAlaIle 144
Db      1213 -----GACCTTATCAGCAAGGCTATTGAATTCACATTAATGATTAATGATTAATG 1263
Qy      145 GluAsnMetAlaTyrlleTyTyrlHisGln-----TyrGluAlaAlaLysSerProThr 162
Db      1264 GAAACCATGGGAACTATTGAAGTACAAAGAGAAACATGAGAAAGCCATTGACATGTTTC 1323
Qy      163 LysAspAspTyTyrAsnAlaLysSerAlaLeuGluArgAlaLeuIle----- 178
Db      1324 AACAAAGCTATTAACTGCGCCAAATCGGAATGAGATGAGCCCATCTGTATTCATTGTC 1383
Qy      179 ---SerGlyThrgLinhIsAspGluIleLysLysSerTyTyr 190
Db      1384 GATGCCGCCCATGCCCAGACAGAAAGTTGCAAGAAATATAC 1422

```

Search completed: July 4, 2003, 05:36:16
Job time : 3200 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 4, 2003, 03:02:42 ; Search time 1085 Seconds
(without alignments)
3045.050 Million cell updates/sec

Title: US-10-069-544-2
Perfect score: 1043
Sequence: 1 MKIRKPMVMAMGLVLSAC.....EIKSYDKLLSDYKLLSDPK 204

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/sgn2.1/USPRO_epool/US10069544/runat_30062003_091537_26270/app_query.fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -DOOPCL=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10069544 @GCN 1.1 1456 @runat_30062003_091537_26270 -NCPU=6 -ICPU=3
-NO MMAP -LARGESUBSTR -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THRAID=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_eestba:*
2: em_eesthum:*
3: em_eestln:*
4: em_eestmu:*
5: em_eestov:*
6: em_eestpl:*
7: em_eestro:*
8: em_hnc:*
9: gb_eest1:*
10: gb_eest2:*
11: gb_hnc:*
12: gb_eest3:*
13: gb_eest4:*
14: gb_eest5:*
15: em_eestfun:*
16: em_eestum:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_fod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	159.5	15.3	684	17 BH389242	BH389242 AG-ND-101
2	122.5	11.7	599	13 BU122799	BU122799 BU122799
3	119.5	11.5	702	13 BU002833	BU002833 BU002833
4	118.5	11.4	582	9 A1067887	A1067887 EST209547
5	115.5	11.1	639	14 BQ087513	BQ087513 Cri-9 B15
6	112	10.7	570	17 BQ203584	BQ203584 Sm1-45C24
7	110.5	10.6	597	10 BE495586	BE495586 WHE09652 G
8	109.5	10.5	603	14 BM736285	BM736285 952057D09
9	105	10.1	499	9 A1067720	A1067720 EST209403
10	104.5	10.0	413	9 A1068135	A1068135 EST209825
11	103	9.9	653	17 BH654661	BH654661 BOMW30TF
12	101.5	9.7	615	10 AM308911	AM308911 sf91f10.Y
13	100.5	9.6	975	17 CNS06N4	AL406514 T3 end of
14	99	9.5	783	9 AL036817	AL036817 DKF2P564L
15	99	9.5	985	17 CNS06J7M	AL401240 T3 end of
16	98.5	9.4	650	17 B88438	B88438 CPG0195A Cp
17	98.5	9.4	858	9 AJ456016	AJ456016 t246e09.Y
18	98.5	9.4	532	10 BE048148	BE048148 t246e09.Y
19	98	9.4	660	12 BF792441	BF792441 602253522
20	98	9.4	866	12 BE910866	BE910866 601661940
21	97.5	9.3	866	9 AU130818	AU130818 AU130818
22	97	9.3	605	14 BQ299342	BQ299342 sa046d01.Y
23	96.5	9.3	671	14 BQ870586	BQ870586 QCD9H11.Y
24	96.5	9.3	327	14 D34850	D34850 CELK013C2F
25	95.5	9.2	1014	17 CNS07DGB	AL440433 T3 end of
26	95.5	9.2	917	13 BG964234	BG964234 602829021
27	95	9.1	717	13 CNS0756E	AL429708 clone BAO
28	95	9.1	919	14 BQ434018	BQ434018 AGENCOURT
29	95	9.1	2062	14 AK005440	AK005440 Mus muscu
30	95	9.1	600	13 B1422511	B1422511 EST533177
31	94.5	9.1	651	13 BU131003	BU131003 BU131003
32	94.5	9.1	651	13 CNS06P2E	AL408828 T3 end of
33	94.5	9.1	1001	14 BQ070967	BQ070967 AGENCOURT
34	94.5	9.1	1015	14 BQ070453	BQ070453 AGENCOURT
35	94.5	9.1	1051	17 CNS06M7H	AL405123 T3 end of
36	94.5	9.1	468	14 BM952706	BM952706 952057D09
37	94	9.0	557	13 BM085569	BM085569 ga126f05.
38	94	9.0	490	13 BM111682	BM111682 EST559218
39	93.5	9.0	587	10 AM622449	AM622449 EST513237
40	93.5	9.0	721	14 BP024656	BP024656 BP024656
41	93.5	9.0	574	13 BM031337	BM031337 496708 MA
42	93	8.9	574	13 BM174957	BM174957 h1pPO_91
43	92.5	8.9	459	13 BQ356760	BQ356760 ILS-HT073
44	92.5	8.9	467	14 B88532	B88532 CPG0196A Cp
45	92.5	8.9	658	17 B88532	

ALIGNMENTS

RESULT 1
BH389242
LOCUS
DEFINITION AG-ND-10107.TR ND-TAM Anopheles gambiae genomic clone AG-ND-10107,
DNA sequence.
ACCESSION BH389242
VERSION BH389242.1 GI:17335383
KEYWORDS
SOURCE
ORGANISM
GSS.
Anopheles gambiae
African malaria mosquito.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE
1 (bases 1 to 684)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submission of BAC-end sequences from *Anopheles gambiae*
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: AG-ND-10107.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@igr.org

This clone is from an *A. gambiae* BAC library (ND-TAM) provided by P.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.
 Location/Qualifiers
 1..684
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-10107"
 /clone_1ib="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 230 a 111 c 145 g 198 t

ALIGNMENT SCORES:

Pred. No.: 5,51e-11 Length: 684
 Score: 159.50 Matches: 50
 Percent Similarity: 47.73% Conservative: 34
 Best Local Similarity: 28.41% Mismatches: 76
 Query Match: 15.29% Indels: 16
 DB: 17 Gaps: 3

US-10-069-544-2 (1-204) x BH389242 (1-684)

QY 9 MelvalMetAlaMetGly-----leuValleuSerAlaCysGlnSerThr----- 23
 DB 82 GTAATATTCGCAATTCGTTATCTGCATTCGTCGACAGGCTGTCAACGACAAATTC 141
 QY 24 ---ProileProProlysaenAnProGlnleuAlaGlnleuArgThrglnleuAlaile 42
 DB 142 GACACATATATAAAGAGTCTGAAAGTCTTTTATGATCTAGCTGCTGCTGCT 201
 QY 43 SerleuLeuAspMetGlyLysleuAspGlnAlaLysGlnleuAspAlaaleuSer 62
 DB 202 GAATATATTCGTCAGGTGATTTAGATGCTGCAAAAGCGCCTAGACCAAGCTTTAGAA 261
 QY 63 AlaAspArgGlnleuAlaProAlaIyArgThrglnleuAlaLysValTyrglnAlaSerGlu 82
 DB 262 ACTGATTCACGTAATTCGCGCAATATGATATGATGCTCTTACTGCAACAGAGGC 321
 QY 83 AspAlaThrHisGlnThrLysAlaGlnArgleuPheGlnLysAlaileGlnleuAsnPro 102
 DB 322 AGTAAGATCAGTATGAAAGACAGCGGTATTTAAGCGTCAATTCACAGTCACTCA 381
 QY 103 LysAspMetGlnSerTyrrhetaAspTyrglyPheTyrglnleuValGlnMetGlyAspLeuSer 122
 DB 382 AAAAATGACCACTGCTGTAATATATGAAACATATTTATTTCTGAGCGCTTAAT 441
 QY 123 GlyAlaLeuLeuTyrrheAspLysProSerArgAlaileGly-TyrglnLysArgVala 142
 DB 442 GACGCGATTAACAATTGAGTATAGCTGCGCTGCGCTTGAATGATACAGCTTACAA 501
 QY 142 AlaileGlnAspMetAlaTyrrileTyrrHisGlnTyrglnAlaLysSerProth 162
 DB 502 GCGTTTAAAGAAATTTAGTGGGTTTACTTAAGTGGTGAATGCTGCA----- 550

QY 162 rlyAspAspTyrrhetaAspAlaLysSerAlaileGlnLysArgAlaileu 177
 DB 551 -----AATGCTGAAAGAAAGCAATTCACCAACAGCTTAA 580

RESULT 2
 LOCUS B0122799 599 bp mRNA linear EST 23-JAN-2002
 DEFINITION B0122799 unpublished oligo-capped cDNA library, C. elegans L1 stage
 ACCESSION B0122799
 VERSION B0122799.1 GI:18282938
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea
 ; Rhabditidae; Pelodierinae; Caenorhabditis.
 1 (bases 1 to 599)
 Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
 and Sugano, S.
 A complementary view of the C. elegans genome
 Unpublished (2002)
 CONTACT Tadasu Shin-i
 CENTER For Genetic Resource Information
 NATIONAL INSTITUTE OF GENETICS
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 TEL: 81-559-81-6856
 FAX: 81-559-81-6855
 EMAIL: tshin@genes.nig.ac.jp.
 LOCATION/Qualifiers
 1..599
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="YK1292B08"
 /clone_1ib="unpublished oligo-capped cDNA library, C. elegans L1 stage"
 /sex="hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="L1"
 /note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

ALIGNMENT SCORES:

Pred. No.: 5,21e-06 Length: 599
 Score: 122.50 Matches: 37
 Percent Similarity: 47.23% Conservative: 24
 Best Local Similarity: 28.68% Mismatches: 55
 Query Match: 11.74% Indels: 13
 DB: 13 Gaps: 4

US-10-069-544-2 (1-204) x B0122799 (1-599)

QY 30 AenProGlnleuAlaGlnleuArgThrglnleuAlaileSerleuLeuAspMetGlyLys 49
 DB 204 AATCTGATTTGATATGCTGCAAGATGATCTTGAAACTTACTTAAGCAATGGGAAGA 263
 QY 50 LeuAspGlnAlaLysGlnleuAspAlaileuSerAlaAspArgGlnPheAlaPro 69
 DB 264 CTGGAAGAGCAAGAGTTGTTACTTGAAGCAATCGAAATCAACACAGTTCGCTGC 323
 QY 70 AlaTyrrArgThrglnleuAlaLysValTyrglnAlaSerGluAsp-----AlaThrHis 86
 DB 324 GATGCTCAATCTTGATGATGATCAATATGCAAGAGAAATTTGTTGCAATTCAT 383
 QY 87 GlnThrLysAlaGlnArgleuPheGlnLysAlaileGlnleuAsnProLysAspMetGln 106
 DB 384 CAT-----TTGAGAAAGCTGTTACTTGTGATCCAAACTTCTCGAC 425

Oy 107 serTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
 Db 426 GCTATATAAATCTTGGAAATGTTCTGAAGAAGCGCAGAGATTTTCATAGAGCGGTTTCA 485
 Oy 127 TyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArgValAlaIleGlu--- 145
 Db 486 GCTTAT-----CTCCGAGCCTTGTAATCTGTCTGTAATCATGACAGTTGTCATGG 536
 Oy 146 AsnMetAlaTyrIleTyrTyrThrIleGln 154
 Db 537 AATTGGCATGTGTACTACGACAG 563
 RESULT 3
 LOCUS BU002833 702 bp mRNA linear EST 05-DEC-2001
 DEFINITION BU002833 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA039602 5',
 mRNA sequence.
 ACCESSION BU002833
 VERSION BU002833.1 GI:17355244
 KEYWORDS EST.
 SOURCE Japanese medaka.
 ORGANISM Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
 Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
 1 (bases 1 to 702)
 Kohara, Y., Shin-I, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.
 Medaka EST Project in Takeda's lab
 Unpublished (2001)
 Contact: Tadao Shin-I
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1..702
 /organism="Oryzias latipes"
 /strain="Hd-rf"
 /db_xref="taxon:8090"
 /clone="MF01SSA039602"
 /clone_1db="MF01SSA cDNA"
 /sex="mixture of female and male"
 /tissue_type="whole embryo"
 /dev_stage="segmentation stage 20 - 25"
 BASE COUNT 183 a 177 c 177 g 165 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.75e-05 Length: 702
 Score: 119.50 Matches: 55
 Percent Similarity: 42.58% Conservative: 34
 Best Local Similarity: 26.32% Mismatches: 75
 Query Match: 11.46% Indels: 45
 DB: 13 Gaps: 10
 US-10-069-544-2 (1-204) x BU002833 (1-702)
 Oy 10 ValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProProLysAsn 29
 Db 68 CTGTGAGCTCGACGAGACATGAGGACAGCTCAAGCTTATGTCTGCTTACAAATAC 127
 Oy 30 AsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
 Db 128 AACCCAGATCTCTATGTGTACGTACGATTTCGGCACTTGCTCAAAAGCCCTGGAGACT 187
 Oy 50 LeuAspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
 Db 188 TTGAGAGGCTTAAGGCTTTGATCCTGAAGCCATGTAGACTACGCCCACTTGTCTGTC 247

Oy 70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAsp-----AlaThrHis 86
 Db 248 GCTTGGAGCAACCTCGGTGTGTGTGTCAATGCCAGGAGGAAATATGCTCGTATACAC 307
 Oy 87 GlnThrLysAlaGlnIleArgPheGlnLysAlaIleGluLeuAsnProLysAspMetGln 106
 Db 308 CAC-----TTTGAAAGGCACTGACATTTGACCCAAATTTCTCGAT 349
 Oy 107 serTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
 Db 350 GCTTATATCAATTTTGAATGTTTGAAGAA-----GCTCGCATC 391
 Oy 127 TyrPheAspLysProSerArgAlaIle---GlyTyr-----Glu 138
 Db 392 TTT-----GACAGACGCTGTGCTGAGTACCTCGAGCTGAGTCTAAGCCCC 439
 Oy 139 GlyArgValAlaAlaIleGluAsnMetAlaTyrIleTyrThrGln----- 154
 Db 440 AACCATGACAGTGTTCACGGGACCTGCTGCTTACTATGAAACAGGACATCATGCAC 499
 Oy 155 -----TyrGluAlaAla-----LysSerProThrLysAspAspTyr 166
 Db 500 CTGGCATTCACACCTACCGGCGGCACTTGAACTGCAGCCCACTTCCAGATCCCTAC 559
 Oy 167 AsnAsnAlaLysSerAlaLeu---GluArgAlaLeuIleSerGlyThrGlnHisAspGlu 185
 Db 560 TGTATCTGGCAACGATGATGAGAGAAAGAAATGTGCT-----GAA 604
 Oy 186 IleLysLysSerTyrAspLysLeuLeu 194
 Db 605 GCAGAAAGTGTCTACACACAGCCTTA 631
 RESULT 4
 LOCUS A1067887 582 bp mRNA linear EST 03-AUG-1998
 DEFINITION EST209574 Schistosoma mansoni, Phil Loverde/Jo Merrick Schistosoma
 mansoni cDNA clone SMNCG34 5' end similar to O-linked GlcNAc
 transferase, p110 subunit, mRNA sequence.
 ACCESSION A1067887
 VERSION A1067887.1 GI:3385854
 KEYWORDS EST.
 SOURCE Schistosoma mansoni.
 ORGANISM Schistosoma mansoni
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 Strigoidae; Schistosomatidae; Schistosomatidae; Schistosoma.
 1 (bases 1 to 582)
 Merrick, J.M., Osman, A., Loverde, P.T., Chandra, I., Glodek, A., Fraser,
 C.M. and Lee, N.H.
 Generation of a Schistosoma sp. EST (SMEST) Catalog & Schistosoma
 Gene Index
 Unpublished (1998)
 Contact: Merrick, J.
 State University of New York at Buffalo
 Department of Microbiology, School of Medicine and Biomedical
 Sciences, Buffalo, NY 14214, USA
 Tel: (716) 829-2158
 Fax: (716) 829-2177
 Email: merrick@acu.buffalo.edu
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..582
 /organism="Schistosoma mansoni"
 /db_xref="taxon:6183"
 /clone="SMNCG34"
 /clone_1db="Schistosoma mansoni, Phil Loverde/Jo Merrick"
 /note="Vector: pBluescript SK; Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT 168 a 119 c 139 g 156 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.75e-05 Length: 582
 Score: 116.50 Matches: 42

Percent Similarity: 44.87% Conservative: 28
 Best Local Similarity: 26.92% Mismatches: 59
 Query Match: 11.36% Indels: 27
 DB: 9 Gaps: 5

US-10-069-544-2 (1-204) x A1067887 (1-582)

QY 10 ValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProProLysAsn 29
 Db 108 TTAGTTGGTGGAGGCGATGATGAGTCTGCAATTAATGCTTATGCAACAGCAGCTGCAATAC 167
 QY 30 AsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
 Db 168 AACCCAGATTATATTTGCGTTCGAAAGTATCTGGGAACCTTAAAGCAGCTTGGACGA 227
 QY 50 LeuAspGlnAlaLysGlnIleLeuAspAlaAlaLeuSerAlaAspArgGlnIleAlaPro 69
 Db 228 TTGGATGAGGGCAATCATGCTATCTGAAGGATTTGAGACCTGTCACCAATTCGACAGT 287
 QY 70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAsp-----AlaThrHis 86
 Db 288 GCTTGGAGCAACCTGGATGTGTTCAATGCTCAAAATGAGATATGTTGGCAATTCAT 347
 QY 87 GlnThrLysAlaGlnArgLeuPheGluLysAlaIleGluLeuAsnProLysAspMetGln 106
 Db 348 CAC-----TTGAAAGGCTGTTGATCTTGGATCTTCTGAT 389
 QY 107 SerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
 Db 390 GCTTACGCTACACCTGGAAACGCTGGAAGAA-----GCTGGCAAT 431
 QY 127 TyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArgValAlaIleGlu--- 145
 Db 432 TTC-----GACCGTGCAAGTACGACGCTTACCTTACGCTTACACTATCTCCA 479
 QY 146 -----AsnMetAlaTyrIleTyrThrIleGln 154
 Db 480 AATAACGCTGTTGTCATGGAAATTTAGCCTGTTTATTATGAGCA 527

RESULT 5 BQ087513 639 bp mRNA linear EST 05-APR-2002
 LOCUS BQ087513
 DEFINITION Crl-9_E15_Sp6 Ceratopteris Spore Library Ceratopteris richardii
 ACCESSION BQ087513.1 GI:20046714
 VERSION BQ087513.1
 KEYWORDS EST.
 SOURCE Ceratopteris richardii.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Filicopsida; Filicales; Pteridaceae; Ceratopteris.
 REFERENCE 1 (bases 1 to 639)
 AUTHORS Chatterjee, A., San Miguel, P., Stout, S.C., Banks, J. and Roux, S.J.
 TITLE Expressed sequence tags of cDNA clones from a C. richardii library.
 JOURNAL Unpublished (2000)
 COMMENT Contact: Roux, SJ
 Section of Molecular Cell and Developmental Biology
 University of Texas
 Biology Building, Room 16, Austin, TX 78712, USA
 Tel: 512 471 4238
 Fax: 512 232 3402
 Email: sroux@uts.cc.utexas.edu
 Place: Crl-9_E15 row: E column: 15
 Seq primer: Sp6.
 Location/Qualifiers
 1. 639
 /organism="Ceratopteris richardii"
 /cultivar="Progn"
 /db_xref="taxon:49495"
 /clone="Crl-9_E15"
 /clone_lib="Ceratopteris Spore Library"
 /tissue_type="Gametophyte"
 /cell_type="Spore"

/dev_stage="20 hours after germination initiation"
 /note="Vector: pCMVSPORT6; EST sequence from cDNA library.
 cDNA library constructed from mRNA isolated from C.
 richardii spores that had developed for 20 hours after
 their germination had been initiated by white light."

BASE COUNT 191 a 141 c 153 g 164 t

ALIGNMENT SCORES:

Pred. No.: 5,29e-05 Length: 639
 Score: 115.50 Matches: 50
 Percent Similarity: 39.69% Conservative: 27
 Best Local Similarity: 25.77% Mismatches: 86
 Query Match: 11.07% Indels: 31
 DB: 14 Gaps: 6

US-10-069-544-2 (1-204) x BQ087513 (1-639)

QY 4 ArgVallySTPProMetVal-----MetAlaMetGlyLeuValLeuSerAla 19
 Db 4 CGCGTCGCTGCTCTAATCTGGCAATGATGAGAAAGGATGATGAGGCT 63
 QY 20 CysGlnSerThrProIleProProLysAsnAsnProGlnLeuAlaGlnIleArgThrGln 39
 Db 64 GAGGAATGTCAGACAAAGCACTCTCAATTAACCTTAAGCTTGATGCTTACAGTAT 123
 QY 40 IleAlaIleSerLeuLeuAspMetGlyLysLeuAspGlnAlaLysGlnIleLeuAspAla 59
 Db 124 CTCGGCAATTTGCTCAAGGTCGCGGTGGACAAACATGCTTATCTGCTACAGTAA 183
 QY 60 AlaLeuSerAlaAspArgGlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGln 79
 Db 184 GCTTGGCCATTCACCTACTTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243
 QY 80 AlaSerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGluLysAlaIleGlu 99
 Db 244 GAAGAGGAGGAC-----TACAGAAAGCTCTCTCTACTTACAGAAAGATCCGG 294
 QY 100 LeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGly 119
 Db 295 CTGAAGCCCACTTCGAGATGACACATGATTAAGAAACGTTCTTAAGCTGTAGGA 354
 QY 120 AspLeuSerGlyAlaLeuIleTyrPheAspLys-----ProSerArgAla 134
 Db 355 CGGCTCAGATGCAATCTTGTTCACGAAGCAATCAATCAGCCCTGATCATCT 414
 QY 135 IleGlyTyrGluGlyArgValAlaAlaIleGluAsnMetAlaTyrIleTyrThrGln 154
 Db 415 GTTGGCTAT-----GTTAATATGGCTAGTCTTACTATGAGCAG 453
 QY 155 -----TyrGluAlaAla-----LysSerProThr 162
 Db 454 AACCACTTGAGCTGGCCATATGAACTATAGCAAGGATCTTAATAGATTCAGCTTTC 513
 QY 163 LysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAla 176
 Db 514 CTAGAGCATACACAACTTGTGTAATCATTTGAAAGATGCA 555

RESULT 6 BH203584 570 bp DNA linear GSS 24-OCT-2001
 LOCUS BH203584
 DEFINITION Sm1-45C24.TF Sm1 Schistosoma mansoni genomic clone Sm1-45C24, DNA
 sequence.
 ACCESSION BH203584.1 GI:16377288
 VERSION BH203584.1
 KEYWORDS GSS.
 SOURCE Schistosoma mansoni.
 ORGANISM Schistosoma mansoni.
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 Strigoidae; Metacercariae; Schistosomatidae; Schistosomidae; Schistosoma.
 REFERENCE 1 (bases 1 to 570)
 AUTHORS Shetty, J., Simpson, A., Malek, J., Koo, H., Loverde, P.T. and El-Sayed,
 N.M.

Db 56 AACCACTGGTGTGATGACATAGTATCTGGGAATCTATGAAAGCCAGGCTCT 115
 Oy 50 LeuapGlnAlaValGlnGlnInleuaspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
 Db 116 GTGCAAGAGGCGATAGCTTGTCTGATGAGCAATACGATAGATCCACACTTGTCAATT 175
 Oy 70 AlaTyrArgThrLeuAlaValValTyrGlnAlaSerGlnAspAlaThrHisGlnThrLys 89
 Db 176 GCATGCTCAAAATCTAGCCCGCTGTTTATGAGAGTGGACCTC-----AATAAG 226
 Oy 90 AlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMet 109
 Db 227 GCTATGCGATATTAACAGAGAGCTGTTAACTTAAGCCATCTTGTGCTGATGACATCTT 286
 Oy 110 AspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAsp 129
 Db 287 AATCAAGGGAATGTGTACAAAGGCTATGGGAATGCTCAGAGGAGCACTTCTGTTATACG 346
 Oy 130 -----LysProSerArgAlaIleGlyTyrGlnGlyArgValAlaAlaIle 144
 Db 347 CGTGTCTTTCAGGACGCGCTTCTAGTACGCCATGCTTAC----- 385
 Oy 145 GluAsnMetAlaTyrIleTyrTyr-----HisGlnTyr 155
 Db 386 GGTATCTTGGCCACTATCTATCTATGAGCAAGCCAGCTGATATGSCATTCATTGTTAC 445
 Oy 156 GluAlaAla-----LysSerProThrLysAspAspTyrAsnAsnAlaLysSerAla 172
 Db 446 AATCAGGCTATACTGTGTGATTTCTCGTTTGTGAGACATACATTAACATGAGCAATGCA 505
 Oy 173 LeuGluArgAla 176
 Db 506 CTTAAAGATGCT 517
 RESULT 8 603 bp mRNA linear EST 01-MAR-2002
 LOCUS BM736285
 DEFINITION 952057D09.x3 952 - BMS tissue from Waldbot Lab (reduced rRNA) Zea
 maays cDNA, mRNA sequence.
 ACCESSION BM736285
 VERSION BM736285.1 GI:19057618
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 603)
 REFERENCE
 AUTHORS Waldbot V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 UNIVERSITY
 JOURNAL Unpublished (1999)
 COMMENT Contact: Waldbot V.
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: waldbot@stanford.edu
 Plate: 952057 row: D column: 09.
 FEATURES
 source
 1. 603
 /organism="Zea mays"
 /cultivar="BMS (Black Mexican Sweet)"
 /db_xref="taxon:4577"
 /clone_id="952 - BMS tissue from Waldbot Lab (reduced rRNA
)"
 /tissue_type="suspension culture"
 /dev_stage="mixed logarithmic and stationary growth
 phases"
 /lab_host="DH10B"
 /note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The
 library was prepared by George Rudenko using poly (A)

selected RNA and Universal Riboclone cDNA Synthesis System
 (Promega). cDNA was synthesized using both random and
 oligo(dT) primers in separate reactions and equipped with
 EcoRI adaptors. Library was size-fractionated on agarose
 gels (for insert size >400bp) and non-directionally cloned
 into EcoRI-digested pUC19 vector. Blue/white selection on
 carbenicillin-containing plates was used to recover
 positive clones."

BASE COUNT 172 a 117 c 133 g 180 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.000319 Length: 603
 Score: 109.50 Matches: 48
 Percent Similarity: 41.86% Conservative: 24
 Best Local Similarity: 27.91% Mismatches: 69
 Query Match: 10.50% Indels: 31
 DB: 14 Gaps: 8

US-10-069-544-2 (1-204) x BM736285 (1-603)

Oy 30 AsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
 Db 93 AATCTCGCTGTGGGAGAGCATGCAACATTTAGAGCTATATCAAAAGAGGATTAAC 152
 Oy 50 LeuaspGlnAlaValGlnGlnInleuaspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
 Db 153 CTTGCAAGAGCTGTGATGATGTATCAAAATGSCACTGTCAATTAAGCCACTTGTGCTCAG 212
 Oy 70 AlaTyrArgThrLeuAlaValTyr-----GlnAlaSerGlnAspAlaThrHisGln 87
 Db 213 TCGCTAAATTAACCTTGAGTGTCTATATCTGTTCAAGGTAAAGATGATTC----- 263
 Oy 88 ThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSer 107
 Db 264 -----GCTGCTACATATTGAGAGGCGCATACATGCAAAATCCACATATGCTGAACA 317
 Oy 108 TyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeu----- 125
 Db 318 TTTATATATCTAGAGTGTTCTTTACAGAGATGCTGTAATATTAATTGACATTCATGCT 377
 Oy 126 -----IleTyrPheAspLysProSerArgAlaIleGlyTyrGlnGlyArgVal 141
 Db 378 TACGAGATGCGCTTCAGATGATCTGATTCAGCAAAATGCTGGC-----CAAGACCGTTTG 434
 Oy 142 ValAlaIleGlnAsnMetAlaTyrIleTyrThrHisGln-----TyrGlnAla 157
 Db 435 CTTGCA-----ATGAACTATATTGACAGAGGCTCAGATGACAACTTTATGAAGCT 485
 Oy 158 AlaLysSer-----ProThrLysAspAspTyrAsnAsn 168
 Db 486 CACAGGAGGTGGGAAAACGCTTTATGAATTGTATTCACAATATACAGTTGGGATTAAC 545
 Oy 169 AlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
 Db 546 TCAAAAGTTGCT---GACCGCCCATGTGATTCGGT 578
 RESULT 9 499 bp mRNA linear EST 03-AUG-1998
 LOCUS A1067720
 DEFINITION EST009403 Schistosoma mansoni, Phil Loyerde/Joel Merrick Schistosoma
 mansoni cDNA clone SMNCD35 5' end similar to O-linked GlcNAc
 transferase, p110 subunit, mRNA sequence.
 ACCESSION A1067720
 VERSION A1067720
 KEYWORDS EST.
 SOURCE Schistosoma mansoni.
 ORGANISM Schistosoma mansoni.
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 Strigeiida; Schistosomatidae; Schistosomatidae; Schistosoma.
 1 (bases 1 to 499)
 REFERENCE Merrick, J.M., Osman, A., Loyerde, P.T., Chandra, I., Glodek, A., Fraser
 , C.M. and Lee, N.H.

TITLE Generation of a Schistosoma sp. EST (SMEST) Catalog & Schistosome Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Contact: Merrick, J.
 State University of New York at Buffalo
 Department of Microbiology, School of Medicine and Biomedical Sciences, Buffalo, NY 14214, USA
 Tel: (716)-829-2158
 Fax: (716)-829-2177
 Email: merrick@acu.buffalo.edu
 Seq primer: M13 Reverse

FEATURES

source

1. 499
 /organism="Schistosoma mansoni"
 /db_xref="taxon:6183"
 /clone="SMNCD35"
 /clone_lib="Schistosoma mansoni, Phil Loverde/Joe Merrick"
 /note="Vector: pBluescript SK; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 144 a 103 c 122 g 130 t

ORIGIN

Alignment Scores:

Pred. No.: 0.0096 Length: 499
 Score: 105.00 Matches: 36
 Percent Similarity: 46.27% Conservative: 26
 Best Local Similarity: 26.87% Mismatches: 58
 Query Match: 10.07% Indels: 14
 DB: 9 Gaps: 3

US-10-069-544-2 (1-204) x A1067720 (1-499)

QY 10 ValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProProLysAsn 29
 DB 108 TTAGTTGCTGCAGGCGATATGAGTCTGCAATTAATGCTTAATGCAACAGCACTGCAGTAC 167
 QY 30 AaAProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
 DB 168 AACCCAGATTTATATGCTTCGAGTATGCTGGAAACCTTAATTAACCACTTGGACGA 227
 QY 50 LeuAaPrglnAlaLysGlnIleuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
 DB 228 TTGGATGAGCGCAATCATCTATCTGAAAGCATGAGACCTGTCCAACTTCCGACGTG 287
 QY 70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAsp-----AlaThrHis 86
 DB 288 GCTTGAGCAACCTGGAGATGTGTTCATCTCAATGCTCAAAATGACATATGCTTGCATCAT 347
 QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGln 106
 DB 348 CAC-----TTTGAAGAGCGCTTACTTGGATCTTCACTTCTTGAT 389
 QY 107 SerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIle 126
 DB 390 GCTTACGTCAACCTCGAAACGTGCTGAAGAAGCTCGCATATTCGACCGTCAGTAGCA 449
 QY 127 TyrPheAspLys-----ProSerArgAlaIle 135
 DB 450 GCCTACCTTAAGCGCTTGAACATATCTCCAAATTAACCGCTGT 491

RESULT 10
 A1068135 413 bp mRNA linear EST 03-AUG-1998
 LOCUS EST009825 Schistosoma mansoni, Phil Loverde/Joe Merrick Schistosoma
 DEFINITION mansoni cDNA clone SMCNCP30 5' end similar to O-linked GlcNAc
 transferase, p110 subunit, mRNA sequence.

ACCESSION A1068135 GI:3386102
 VERSION A1068135
 KEYWORDS EST,
 Schistosoma mansoni,
 Schistosoma mansoni,
 Eukaryota, Metazoa, Platyhelminthes, Trematoda, Digenea,
 Strigidae, Schistosomatidae, Schistosomatidae, Schistosoma.

REFERENCE 1 (bases 1 to 413)
 AUTHORS Merrick,J.M., Oaman,A., Loverde,P.T., Chandra,I., Glodex,A., Fraser
 C.M. and Lee,N.H.
 TITLE Generation of a Schistosoma sp. EST (SMEST) Catalog & Schistosome
 JOURNAL Unpublished (1998)
 COMMENT Contact: Merrick, J.
 State University of New York at Buffalo
 Department of Microbiology, School of Medicine and Biomedical Sciences, Buffalo, NY 14214, USA
 Tel: (716)-829-2158
 Fax: (716)-829-2177
 Email: merrick@acu.buffalo.edu
 Seq primer: M13 Reverse

FEATURES

source

1. 413
 /organism="Schistosoma mansoni"
 /db_xref="taxon:6183"
 /clone="SMNCP30"
 /clone_lib="Schistosoma mansoni, Phil Loverde/Joe Merrick"
 /note="Vector: pBluescript SK; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 120 a 82 c 103 g 108 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.00818 Length: 413
 Score: 104.50 Matches: 31
 Percent Similarity: 49.06% Conservative: 21
 Best Local Similarity: 29.25% Mismatches: 45
 Query Match: 10.02% Indels: 9
 DB: 9 Gaps: 2

US-10-069-544-2 (1-204) x A1068135 (1-413)

QY 10 ValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProProLysAsn 29
 DB 108 TTAGTTGCTGCAGGCGATATGAGTCTGCAATTAATGCTTAATGCAACAGCACTGCAGTAC 167
 QY 30 AaAProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
 DB 168 AACCCAGATTTATATGCTTCGAGTATGCTGGAAACCTTAATTAACCACTTGGACGA 227
 QY 50 LeuAaPrglnAlaLysGlnIleuAspAlaAlaLeuSerAlaAspArgGlnPheAlaPro 69
 DB 228 TTGGATGAGCGCAATCATCTATCTGAAAGCATGAGACCTGTCCAACTTGCATCAT 287
 QY 70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAsp-----AlaThrHis 86
 DB 288 GCTTGAGCAACCTGGAGATGTGTTCATCTCAATGCTCAAAATGACATATGCTTGCATCAT 347
 QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGln 106
 DB 348 CAC-----TTTGAAGAGCGCTTACTTGGATCTTCACTTCTTGAT 389
 QY 107 SerTyrMetAspTyrGly 112
 DB 390 GCTTACGTCAACCTCGGA 407

RESULT 11
 BH654661 653 bp DNA linear GSS 19-FEB-2002
 LOCUS BOMMG30TF_BO_2_3_KB Brassica oleracea genomic clone BOMMG30, DNA
 DEFINITION sequence.

ACCESSION BH654661
 VERSION BH654661.1 GI:18712884
 KEYWORDS GSS,
 Brassica oleracea,
 Brassica oleracea,
 Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
 Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots,
 Rosidae, eurosids II, Brassicales, Brassicaceae, Brassica.
 REFERENCE 1 (bases 1 to 653)

AUTHORS Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
 TITLE Whole genome shotgun sequencing of *Brassica oleracea*
 JOURNAL Unpublished (2001)
 COMMENT Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF

Class: sheared ends.

FEATURES
 source

1..653
 /organism="Brassica oleracea"
 /strain="TO100DH3"
 /db_xref="taxon:3712"
 /clone_lib="BOMWG30"
 /clone_lib="BO_2_3_KB"
 /note="Vector: pBOS1, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pBOS1 using BstXI linkers"

BASE COUNT 153 a 158 c 138 g 204 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.00284 Length: 653
 Score: 103.00 Matches: 40
 Percent Similarity: 39.43% Conservative: 29
 Best Local Similarity: 22.86% Mismatches: 68
 Query Match: 9.88% Indels: 38
 DB: 17 Gaps: 6

US-10-069-544-2 (1-204) x BH654661 (1-653)

31 ProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLeu 50
 630 CCAATTTTATTAACCGCTGCTACAGTAAGCATCGTTGCTGTAATAAGCGAAGCAT 571
 51 AspGlnAlaGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAla 70
 570 GAAGAAGCGGTACAGCACTTACGAAAAAGCGGTAAATTAAACAGATTATTATCAAGA 511
 71 TyrArgThrLeuAlaValTyrGlnAlaSerGlnAspAlaThrIleGlnThrLeuAla 90
 510 TGGTATTAATTTAGCGCTGCTGCTAC-----CACGAATGGCTGCTGACACAGCA 460
 91 GlnArgLeuPheGlnIleAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAsp 110
 459 ATTGAGTGTACAGCAAGCGCTTGAATTCGAAAGAAATTCAGCGCTGTACAAAT 400
 111 TyrGly-----PheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPhe 128
 399 AGAGCGAATGCTCAGTACAAATTTGACGCG----- 370
 129 AspLysPheSerArgAlaIleGlyTyrGlnGlyArgValAlaAlaIleGlnMetAla 148
 369 -----TACAGAGATGCCATGCCCTCTTACACAGCAAGCGCTC 334
 149 TyrIleTyrTyrHisGlnTyrGlnAlaAlaLysSerProThrLysAspAspTyrAsnAsn 168
 333 TATGTC-----AAGCGGAGACTATTCGAA 310
 169 AlalysSerAlaLeuGlnAlaLeuIleSerGlyThrGlnHisAspGlnIleLysLys 188
 309 GCTTGTACAGCGCGCAATGCGCTGTAGCGGTAAAGATACGAGATGCGATCGCC 250
 189 SerTyrAspLysLeuLeuSerAspTyrLysLeuLeuSerAspTyr 203
 249 TCTTATGATTAAGCAATTCCG---TACAA-----CCGATTAC 214

RESULT 12
 AM308911 615 bp mRNA linear EST 02-DEC-2001
 LOCUS AM308911

DEFINITION sf91f10.Y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:

Gm-c1019-3524.5' similar to TR:082039 082039 SPINDLY PROTEIN. ;,

mRNA sequence.

ACCESSION AM308911 GI:6724512

VERSION KEYWORDS

SOURCE

ORGANISM

Glycine max
 soybean.
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustroids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

1 (bases 1 to 615)

REFERENCE

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@wustl.edu
 This clone is available through: Reggen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: ccu@reggen.com
 Insert Length: 2128 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 444.

FEATURES
 source

1..615
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1019-3524"
 /clone_lib="Gm-c1019"
 /tissue_type="immature seed coats of greenhouse grown plants"

/lab_host="DH10B (Gibco BRL)"
 /note="Vector: pSPORT1 (Life Technologies); Site 1: Not I; Site 2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies superscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."

BASE COUNT 166 a 112 c 162 g 174 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.00412 Length: 615
 Score: 101.50 Matches: 30
 Percent Similarity: 44.90% Conservative: 14
 Best Local Similarity: 30.61% Mismatches: 51
 Query Match: 9.73% Indels: 3
 DB: 10 Gaps: 2

US-10-069-544-2 (1-204) x AM308911 (1-615)

33 LeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeuAspGln 52
 295 CTTGCAATTTTATTAACAGATTGCTACCAACATTAAGCTGCGAGCAACTCAGAGAT 354

Qy 53 AlAtySGInLLeuAspAlaLeuSerAlaAspArgInpHealAproAlaTyArg 72
 Db 355 GGAATTCAAAATATTATTGAGCCCTCAAGAGTATCCGCACTATGCTCTGCAATGAT 414
 Qy 73 ThrLeuAlaLeuValTyArgInAlaSerGluAspAlaThrHisGlnThrLeuAlaGlnArg 92
 Db 415 AATCTTGCTGTGCTGCTAT-----TCTGAATGATGCAATATGACACA---GCCCTCAGT 465
 Qy 93 LeuPheGluValAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGly 112
 Db 466 TTCTATGAAGAAAGCGTCGATCAGAGAGCGCATCTATCCGAGCTTATTCGAATATGAGCT 525
 Qy 113 PheTyrLeuValAlaMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLys 130
 Db 526 GTGATGTTTAAATCCGCGTATTTGCAATCGCTATTACTTCTTATGAGAGG 579

RESULT 13
 CNS06N4/C 975 bp DNA linear GSS 17-JUN-2001
 LOCUS T3 end of clone AU0AA010F05 of library AU0AA from strain CBS 3082
 DEFINITION of Saccharomyces kluyveri, genomic survey sequence.
 ACCESSION AL406514.1 GI:12170615
 VERSION AL406514.1 GI:12170615
 KEYWORDS GSS.
 SOURCE Saccharomyces kluyveri.
 ORGANISM Saccharomyces kluyveri.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 975)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bojot,F.,Pukhara,M., Bon,E., Brotier,P., Lepingle,A., Llorente,B.,
 de-Montigny,J., Dujon,B., Durans,P., Lepingle,A., Llorente,B.,
 Maupertuy,A., Neugebise,C., Olier-Kalogeropoulos,O., Portier,S.,
 Saudin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)

TITLE
 JOURNAL MEDLINE 20584711
 PUBMED 1152876
 2 (bases 1 to 975)
 Neugebise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
 Galliardin,C. and Casaregola,S.
 Genomic exploration of the hemiascomycetous yeasts: 9.
 Saccharomyces kluyveri
 FEBS Lett. 487 (1), 56-60 (2000)

JOURNAL MEDLINE 20584719
 PUBMED 1152884
 3 (bases 1 to 975)
 Genoscope.
 Direct Submision
 Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 laevis var. laevis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbicola,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES
 SOURCE
 1..975
 /organism="Saccharomyces kluyveri"
 /strain="CBS 3082"
 /db_xref="taxon:4934"
 /clone="AU0AA010F05"
 /clone_11b="AU0AA"
 /note="end : T3"

misc_feature complement(<7..>854)
 /note="similar to Saccharomyces cerevisiae ORF YOR007c [SGR2 : similarity to protein phosphatases] 1 putative frimeshift(s)"
 /evidence=not_experimental

BASE COUNT 229 a 223 c 206 g 316 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.0123 Length: 975
 Score: 100.50 Matches: 38
 Percent Similarity: 40.46% Conservative: 15
 Best Local Similarity: 29.01% Mismatches: 45
 Query Match: 9.64% Indels: 33
 DB: 17 Gaps: 4

US-10-069-544-2 (1-204) x CNS06N4 (1-975)

Qy 65 ArgInPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAspAla 84
 Db 453 CGCCACAGCGCTGCGCGCTATCTCTTGAAGAAATACAGAGAAATCCGTTCAAGATGCT 394
 Qy 85 ThrHisGlnThrLysAlaGlnArgLeuPheGluLysAlaIleGluLeuAsnProLysAsp 104
 Db 393 -----GAGTCTGCTATTATTAAGTGAACCTTCTTAT 364
 Qy 105 MetGlnSerTyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAla 124
 Db 363 TCCAAGGGGATCTCTAGATTGAGATTGCAAGATATGCTTAGGT----- 319
 Qy 125 LeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGluGlyArgValAlaIle 144
 Db 318 -----AGCCGAGAGAGAGCTTAGAGCTTCACAAACGTTTGGACATT 274
 Qy 145 GluAsnMetAlaTyrIleTyrThrHisGlnTyrGluAlaAlaLysSerProThrLysAsp 164
 Db 273 GAGGCG-----GACAAAGCCACTGATGTCATGAAGAGA 241
 Qy 165 AspTyrAsnAlaAlaLysSerAlaLeuGlnArgAlaLeu-----IleSerGly 180
 Db 240 GATTACGAAGACTGCCAGAGAGAGAGTTGACAAATCTTGAACCTTGAGAAAGAGTCCCA 181
 Qy 181 ThrGlnHisAspGluIleLysLysSerTyrAsp 191
 Db 180 ACACAAACCGATGAACAGAGACGACGCGAT 148

RESULT 14
 AL036817 783 bp mRNA linear EST 29-FEB-2000
 LOCUS DKFZP564L0963.r1.564 (synonym: hdbx2) Homo sapiens cDNA clone
 DEFINITION DKFZP564L0963.5', mRNA sequence.
 ACCESSION AL036817
 VERSION AL036817.1 GI:5406318
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 783)
 Duesterhoeft,A., Lauber,J., Mewes,H.W., Gaassenhuber,J. and Wiemann
 S.
 EST (Duesterhoeft, et al.)
 Unpublished (1999)
 CONTACT: Duesterhoeft A
 MFS
 Am Klopferpitz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Oligen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.
 This clone (DKFZP564L0963) is available at the RZPD in Berlin.

Db 557 ATAGCGAATTGCTTCTTACAAAAGCAGCATGCTGCAATTAAGCTTTGAAAA 616
 Qy 64 -----AspArgGlnPheAlaProAlaTyrArgThrLeuAlaValTyrGln 79
 Db 617 GCTACTCATATGATCCAAATTTTGGATATGGCTAT---ACTTTGCAAGGCCATAGCAT 673
 Qy 80 AlaSerGluAspAlaThrHisGlnThrLysAlaGlnArgLeuPheGluAlaIleGlu 99
 Db 674 TCTTCAACGATTCCTTCAGATTCCTGCCAAGACG-----TGTTATAGAAAGGCACTAGCG 727
 Qy 100 LeuAsnProLysAspMetGlnSerTyrMetLeuPheTyrGlyPheTyrLeuValGlnMetGly 119
 Db 728 TGTGATCTCCCAATTAATATGATCATATGCGTTGGGACAGAGCTTATATGAATTAAGCG 787
 Qy 120 AspleuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyrGluGly 139
 Db 788 CAATACGAAGAGCTTATATATTTTGAAGAA---GCAGATTCATCATCTGTCAC 844
 Qy 140 ArgValVal-----AlaIleGluAsnMetAlaTyr----- 149
 Db 845 GTAATTTGATCTGTGCTGTGCTGTGCTTCTTGAAGAAACGATTAATAGAAAGTA 904
 Qy 150 IleTyrTyrHisGlnTyrGluAlaAlaLysSerProThr 162
 Db 905 TTACAGTATATGAACTAGCTTCCAGTTACAAACGACT 943
 RESULT 16
 B88438 650 bp DNA linear GSS 12-MAY-2000
 LOCUS B88438
 DEFINITION CpG195A CpIOMAGDNA1 Cryptosporidium parvum genomic similar to cell division control protein 23 (cdc23) and O-linked GlcNAc transferase
 'DNA sequence.
 ACCESSION B88438
 VERSION B88438.1 GI:2970864
 KEYWORDS GSS.
 SOURCE Cryptosporidium parvum.
 ORGANISM Cryptosporidium parvum
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
 Cryptosporididae; Cryptosporidium.
 1 (bases 1 to 650)
 Strong, W.R. and Nelson, R.G.
 Preliminary profile of the Cryptosporidium parvum genome: an
 expressed sequence tag and genome survey sequence analysis
 Mol. Biochem. Parasitol. 107 (1), 1-32 (2000)
 20183851
 Contact: Nelson, R. G.
 Depts. of Medicine & Pharmaceutical Chemistry
 San Francisco General Hospital-University of California, San
 Francisco
 Box 0811, San Francisco, CA 94143-0811, USA
 Tel: 415 206 8846
 Fax: 415 206 3353
 Email: malari@itsa.ucsf.edu
 Submitted sequence has been edited to remove vector sequences 5' to
 the insert, to correct miscalled bases and assign uncalled (N)
 bases throughout the sequence, and to terminate when base-calling
 became ambiguous.
 Seq primer: M13(-21) FORWARD
 Class: Shotgun
 High quality sequence stop: 650.
 Location/Qualifiers
 1..650
 /organism="Cryptosporidium parvum"
 /strain="IOWA"
 /db_xref="taxon:5807"
 /clone_lib="CpIOMAGDNA1"
 /lab_host="E. coli XL2 Blue MRF"
 /note="Vector: pBluescript II (SK-); Site 1: EcoRV, C.
 parvum (IOWA isolate) genomic DNA was hydrodynamically
 sheared to produce fragments having a tight size
 distribution between 2-4 kb by Dr. Yvonne Thorstenson of
 the Stanford DNA Sequencing and Technology Center
 (http://sequence-www.stanford.edu/group/techdev/shear.htm

). The randomly sheared gDNA was chromatographed on
 Sephacryl S-400 to remove any small fragments and DNA
 eluting in the void volume was subcloned into an EcoR
 V-digested, alkaline phosphatase-treated pBluescript II
 (SK-) vector and transformed into E. coli strain XL2 Blue
 MRF+. Recombinant clones from the first plating of the
 library were selected for sequence analysis using 73 and
 77 primers."
 BASE COUNT 208 a 130 c 125 g 187 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0117 Length: 650
 Score: 98.50 Matches: 45
 Percent Similarity: 38.89% Conservative: 39
 Best Local Similarity: 20.83% Mismatches: 75
 Query Match: 9.44% Indels: 57
 Ds: 17 Gaps: 7
 US-10-069-544-2 (1-204) x B88438 (1-650)
 Qy 6 LysTPRPrometValMetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIle 25
 Db 3 AAATGGAGGAAGCTCTT-----TGCAATATATACCAATA 38
 Qy 26 ProProLysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeu 45
 Db 39 TCCATGAGAGTAAAGATTAACAGTCCGATATTTGAAATAGATTTCCAGATGTAACGA 98
 Qy 46 AspMetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaValLeuSerAlaAspArg 65
 Db 99 GAGCTGGGAGACTTGAAGATCACTTAAGCTCAGATTCATTCATTAGCAAGAACAA 158
 Qy 66 GlnPheAlaProAlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGluAsp----- 83
 Db 159 GCTTACTTGGGAATTCATAGAACTAGCTTCATATATTTGATCAAGAAATATTAA 218
 Qy 83 ----- 83
 Db 219 GAACATAATCTTCTTGCAAAAGGCTGTAGAGCTTCCAAATCTCTGTGATCTTGC 278
 Qy 84 -----AlaThrHisGlnThrLysAlaGlnArgLeuPhe 94
 Db 279 GTTGACTTGGAAATCATATGCTGCTCACTAATGATTAATCAAGCTCAGATTCTAC 338
 Qy 95 GluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyr 114
 Db 339 AAGAGAGCTCTTATGATTCAGAGCAAGCCAACTTGGGCTTGGCGTTATGCT 398
 Qy 115 LeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAla 134
 Db 399 CTTGATGATTAATGAGGATATCAGAGTCCCTTTTGGCTTAC-----AGAACGCC 449
 Qy 135 IleGlyTyrGlu---GlyArgValAlaIleGluAsnMetAlaTyrIleTyr----- 151
 Db 450 TTGAATTCGACCTCCTCAAAATCTCAGGCATTTTGGAAATGCGCAAAATTTATCCAGA 509
 Qy 152 -----TyrHisGlnTyrGluAlaAlaLysSerProThrLysAsp 164
 Db 510 CTTATCTTCAAGCTTACTTACGATTAACCTTACGAAAGGCTCTTAACCAATCTCTCGG 569
 Qy 165 Asp-----TyrAsnAsnAlaLysSerAlaLeuGluArgAla 176
 Db 570 GATGCTTCTTCTTGGTACACCAAGAAATTTCC--TTGAAAAAGCT 614
 RESULT 17
 A1456016 858 bp mRNA linear EST 22-APR-2002
 LOCUS A1456016
 DEFINITION A1456016 riken1 Gallus gallus cdna clone 765r1. mRNA sequence.
 ACCESSION A1456016
 VERSION A1456016.1 GI:20266112
 KEYWORDS EST.
 SOURCE chicken.

Db 160 TATCCCTCTTCTACCCATCTATAGCAAGCAGAG-----AACCAAGCAAGGCA 210
 QY 91 GlnArgLeuPheGluValAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAsp 110
 Db 211 CTTGATGCTATAGACAGGCTCTCCAGCTGAACCAAGACCCAAAGTCAATTTCTGAA 270
 QY 111 TyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLys 130
 Db 271 CTTTTTTC-----ACAAAGGAAACCAATTAGAGACAGACAACTCTGCACAA 321
 QY 131 ProSerArgAlaIleGlyTyrGluGlyArgValValAlaIleGlu----- 145
 Db 332 -----GCTTTGAGAGCTATAGAGGCTGCTGCAATTAACCCAGACCA 366
 QY 146 -----AsnMetAlaTyrIleTyrTyrHisGlnTyrGluAlaAlaLysSer 160
 Db 367 GCACAGCGCTGATGAACATGGGTGGCATCCAAAC----- 402
 QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
 Db 403 ---ATCAAGGGAATAATGCTGCTCAGAGCTTATATAGAGAGACCTTA----- 450
 QY 181 ThrGlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeu 200
 Db 451 -----CAGCTGGTTCCAGACGACCAACTGCTG 477
 QY 201 SerAsp 202
 Db 478 AAGGAA 483
 RESULT 19
 LOCUS BF792441 660 bp mRNA linear EST 12-JAN-2001
 DEFINITION 602253222P1 NIH_MGC_84 Homo sapiens CDNA clone IMAGE:4345877 5',
 mRNA sequence.
 ACCESSION BF792441 GI:12097495
 VERSION BF792441.1 GI:12097495
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 660)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LHAM9966 row: 1 column: 06
 High quality sequence stop: 660.
 Location/Qualifiers
 1. 660
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4345877"
 /clone_1lb="NIH_MGC_84"
 /tissue_type="adrenal cortex carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: adrenal gland; Vector: pCMV-SPORT; Site: 1:
 NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT
 primed. Average insert size 1.229 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC library."
 211 a 155 c 158 g 136 t

Alignment Scores:
 Pred. No.: 0.014 Length: 660
 Score: 98.00 Matches: 43
 Percent Similarity: 39.56% Conservative: 29
 Best Local Similarity: 23.63% Mismatches: 66
 Query Match: 9.40% Indels: 44
 DB: 12 Gaps: 6
 US-10-069-544-2 (1-204) x BF792441 (1-660)
 QY 31 ProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeu 50
 Db 126 CTTCTCGAGAGAGCTCCGCTGGCAGCTGCTCAGGTTTGGCCGTAGGCTGACCA 185
 QY 51 AspGlnAlaLysGlnGlnLeuAspAlaAlaLeuSerAlaAspArgGlnPheAlaProAla 70
 Db 186 AAGAAAGCTGAAAGATACCAATGCACTGTGTGACAGAGACCGGATGCTTGAATGC 245
 QY 71 TyrArgThrLeuAlaLysValTyrGlnAlaSerGluAspAlaThrHisGlnThrLysAla 90
 Db 246 TATGCGCTCTGTGAGCATCTATAGCAAGCAGAG-----AACCAAGCAAGGCA 296
 QY 91 GlnArgLeuPheGluValAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAsp 110
 Db 297 CTTGATGCTATAGACAGGCTCTCCAGCTGAACCAAGGACCCAAAGTCAATTTCTGAA 356
 QY 111 TyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLys 130
 Db 357 CTTTTTTC-----ACAAAGGAAACCAATTAGAGACGACCAACTTCTGCACAA 407
 QY 131 ProSerArgAlaIleGlyTyrGluGlyArgValValAlaIleGlu----- 145
 Db 408 -----GCTTTGAGAGCTATAGAGGCTGCTGCAACTAAACCCAGACCA 452
 QY 146 -----AsnMetAlaTyrIleTyrTyrHisGlnTyrGluAlaAlaLysSer 160
 Db 453 GCACAGCGCTGATGAACATGGGTGGCATCCAAAC----- 488
 QY 161 ProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGluArgAlaLeuIleSerGly 180
 Db 489 ---ATCAAGGGAATAATGCTGCTCAGAGCTTATATAGAGAGCCTTA----- 536
 QY 181 ThrGlnHisAspGluIleLysLysSerTyrAspLysLeuLeuSerAspTyrLysLeuLeu 200
 Db 537 -----CAGCTGGTTCCAGACGACCAACTGCTG 563
 QY 201 SerAsp 202
 Db 564 AAGGAA 569
 RESULT 20
 LOCUS BE910866 964 bp mRNA linear EST 29-SEP-2000
 DEFINITION 601661940P1 NCI_CGAP_Mam1 Mus musculus CDNA clone IMAGE:3962337 5',
 mRNA sequence.
 ACCESSION BE910866
 VERSION BE910866.1 GI:10407850
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 964)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM9128 row: 0 column: 10
 High quality sequence stop: 621.
 Location/Qualifiers

FEATURES

SOURCE

1..964
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3962337"
 /clone_lib="NCI-CCAP Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DHI08"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 262 a 216 c 241 g 245 t
 ORIGIN

Alignment Scores:

Pred. No.: 0.0265 Length: 964
 Score: 98.00 Matches: 50
 Percent Similarity: 40.98% Conservative: 25
 Best Local Similarity: 27.32% Mismatches: 70
 Query Match: 9.40% Indels: 38
 DB: 12 Gaps: 9

US-10-069-544-2 (1-204) x BE910866 (1-964)

QY 29 Aaanaaproglnleuaaglnleargthrglnlealelleserleuenspmetgly 48
 125 AATAAC-----CTTGCCCAACATCAACG-----GACAGGCGC 157
 49 Lysleuaspclnalaalysglnleuaspalaaleuaserlasparglnpheala 68
 158 AACATTGAAGAGGCACTGCGCTGATGCGCAAGCATTTGAAGCTCTCCACAGATTGCT 217
 69 ProalatyrarthrleuualalysValTYrglnalasergluaspaalathrhsglnthr 88
 218 GCGGCACTTCATTTAGCAAGTATGACGCA-----CAGCAGGCGC 259
 89 LysAlaGlnArgleu-----PheGlnLysAlaIleGlnleuasnProlysaspmet 105
 260 AAGCTGCAGAGACACTGATGACATTAAGAGCCATGCAATTGCTCTACATTGCT 319
 106 GlnserTYmetaspTYrglyPheTYrleuValGlnMetGlyaspleserGlyAlaleu 125
 320 GATGCTTATTCATATGGAAACACTCTTAAGAGATGAGATGCGAGGCGCTTGG 379
 126 IleTYrPheaspLysProserArGalaIleGlyTYrGlnGlyArValAlaIle 144
 380 CAGTGTAT-----ACTCGGCCATCCAGATTATCTCTTGTGATGACAC 430
 145 GlnAsnMetAlaTYrIleTYr-----TYR 152
 431 AGCAATCTGGCTCCCATTCACAAAGATTGAGGGAATATCCAGAGCAATAGCTTCTTAC 490
 153 HieGlnTYrGlnAlaIalysSerProThrLysaspAspTYrAsnAlaIalysSerAla 172
 491 CGACAGACTGTAACCTTAAGCTGACTTCCTGATCTTATGTAACCTTGCGCTCATTCG 550
 173 LeuGlnArgAlaLeuIleSerGlyThrhIleAspGlnIleLysLysSerTYrAspLys 192
 551 CTACAG-----ATTGCTGTGATTGACAGACTATGATGAGCGATGAAG-----AAA 598
 193 LeuLeuSer 195
 599 TTGGTTAGT 607
 RESULT 21
 AU130818

LOCUS AU130818 866 bp mRNA linear EST 01-AUG-2002
 DEFINITION AU130818 NT2RP3 Homo sapiens cDNA clone NT2RP3001492 5', mRNA
 sequence.
 ACCESSION AU130818
 VERSION AU130818.1 GI:10991172
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 866)
 Oca,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.
 HRI human cDNA project
 Unpublished (2000)
 CONTACT: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel.: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 Location/Qualifiers

FEATURES

SOURCE

1..866
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RP3001492"
 /clone_lib="NT2RP3"
 /cell_type="teratocarcinoma"
 /note="Vector: pME18SF13; mRNA from NT2 neuronal precursor
 cells after 2-weeks retinoic acid (RA) induction"

BASE COUNT 240 a 184 c 198 g 240 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.0259 Length: 866
 Score: 97.50 Matches: 52
 Percent Similarity: 39.80% Conservative: 26
 Best Local Similarity: 26.53% Mismatches: 83
 Query Match: 9.35% Indels: 36
 DB: 9 Gaps: 8

US-10-069-544-2 (1-204) x AU130818 (1-866)

QY 17 LeuSerAlaCyGlnSerThrProIleProProlysaAaanaPProGlnleuaaglnle 36
 48 CTCTCGCTGTGT-CCGACCAGTCAGACACTCTCGAATAC-----CTAGCCAAATATC 100
 37 ArgThrGlnIleAlaIleSerleuaspmetGlyLysleuaspGlnAlaIalysGln 56
 101 AAACGA-----GACACGGGAAACATTGAAGAGCGAGTTCCTTG 139
 57 LeuaspAlaAlaLeuSerAlaaspArgIlePheAlaProAlaTYrArgThrleuAlaLys 76
 140 TATCGTAAGCATTTGAAGCTTCCCAAGATTTGCTGCTGCCCATTCCAATTTACAGACT 199
 77 ValTYrGlnAlaSerGlnLyspaalathrhsglnthrLysAlaGlnArgleu----- 93
 200 GTRACGCG-----CAGCAGGGAACCTGACGAGAGCTGTGATCAT 241
 94 PheGlnLysAlaIleGlnleuasnProlysaspMetGlnserTYrMetaspTYrGlyPhe 113
 242 TATAGGAGGCTATTGCAATCAGTCTTACCTTGTGATGCTCTACTCTTAATATGGAAC 301
 114 TYrleuValGlnMetGlyAspLeuSerGlyAlaLeuIleTYrPheaspLysProserArg 133
 302 ACTTAAGAGAGATGACAGATGTTTCAGGAGGCTTGCAGCTGTAT-----ACGGCT 352

QY 134 AlailegIYrGLuGIYArGValVal---AlailegluAmMetAlaTYrIleTYr--- 151
 DB 353 GCCATCCAAATTAATCCGATTCGATTCAGATGACATAGCATAGCATTCGCTTCATTCATAG 412
 QY 152 -----TYRHeGInTYrGLuAlaAlaIYsSer 160
 DB 413 GATTCAGGGAATATTTCAGAACCCATAGCTTCTTCCGACCGCTCTGAAACTTAAGCT 472
 QY 161 ProthrIYAspAspTYrAsnAlaIYsSerAlaLeuGIYArGAlaLeuIleSerGIY 180
 DB 473 GATTTCTGATGCTGATTGTAATCTTGCTCATTCGCTGACG---ATTGCTGTGATTGG 529
 QY 181 ThrGlnHIsapGIuIleIYsSerTYrAspIYsLeuIleuSerAsp 196
 DB 530 ACAGACTATGATGACGAATGAAGAAGTTCGACGATTGTCGCTGAC 577

RESULT 22
 CNS06C93/ 982 bp DNA linear GSS 14-JUN-2001
 LOCUS T3 end of clone AROAA003A11 of library AROAA from strain CBS 732 of
 DEFINITION Zygosaccharomyces rouxii, genomic survey sequence.
 ACCESSION AL392221.1 GI:12141366
 VERSION AL392221.1 GI:12141366
 KEYWORDS GSS.
 SOURCE Zygosaccharomyces rouxii.
 ORGANISM Zygosaccharomyces rouxii.
 BUKARYOTA: Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.
 REFERENCE 1 (bases 1 to 982)
 AUTHORS Souciet,J.L., Aligle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S.,
 deMontigny,J., Dujon,B., Durren,P., Lepingle,A., Lloret,B.,
 Malperuy,A., Neugeglise,C., Ozier-Kalogiropoulos,O., Portier,S.,
 Sautin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEMS Lett. 487 (1), 3-12 (2000)

TITLE
 JOURNAL FEMS Lett. 487 (1), 52-55 (2000)
 MEDLINE 20584718
 PUBMED 11152883
 REFERENCE 2 (bases 1 to 982)
 AUTHORS de Montigny,J., Straub,M., Portier,S., Tekala,F., Dujon,B.,
 Wincker,P., Artiguenave,F. and Souciet,J.
 Genomic exploration of the hemiascomycetous yeasts: 8.
 Zygosaccharomyces rouxii
 FEMS Lett. 487 (1), 52-55 (2000)

TITLE
 JOURNAL FEMS Lett. 487 (1), 52-55 (2000)
 MEDLINE 20584718
 PUBMED 11152883
 REFERENCE 3 (bases 1 to 982)
 AUTHORS Genoscope.
 JOURNAL Direct Submission
 SUBMITTED (06-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermocolerans, Kluyveromyces
 laetis var. laetis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia soboloffii,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES
 SOURCE
 1..982
 location/Qualifiers
 /organism="Zygosaccharomyces rouxii"
 /strain="CBS 732"
 /db_xref="taxon:4956"
 /clone="AROAA003A11"
 /clone_1lb="AROAA"
 /note="end : T3"

misc_feature complement (<22..>960)
 /note="similar to Saccharomyces cerevisiae ORF YGR047c [TFC4 ; TFC11c (transcription initiation factor) subunit, 131 kd]"
 /evidence=not experimental

BASE COUNT 251 a 230 c 165 g 335 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.0375 Length: 982
 Score: 97.00 Matches: 43
 Percent Similarity: 42.56% Conservative: 40
 Best Local Similarity: 22.05% Mismatches: 82
 Query Match: 9.30% Indels: 30
 DB: 17 Gaps: 7

US-10-069-544-2 (1-204) x CNS06C93 (1-982)

QY 28 LysAsnAsnProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMet 47
 DB 567 GAGATTTCGACGACATCGAGATTGTAATTCGACGATTCGCTTTA----- 517

QY 48 GlyIYsLeuAspGlnAlaIYsGlnIleuAsp-----AlaIleuSerAlaAspArg 65
 DB 516 TCCAAGGTGAAGAACACAGAACCCATCGATTCCTTACGCCCATCTCGCAATTGAA 457

QY 66 GlnPheAla-----ProAlaTYrArgThrIleuAlaIYsValTYrGlnAlaSerGIYAsp 83
 DB 456 GAGTACTGTAATTCGATTCGATTCGACCATTCGACGATTCGACGATTCGACGATTCGACGAT 397

QY 84 AlaThrHisGlnThrIYsAlaGlnIleArgThrGlnIleuAlaIleGlnLeuAsnProIYs 103
 DB 396 -----TATGATATCCAGCAAGACCTTTATAGAAAGATGATCTTGACCCAAAT 346

QY 104 AspMetGlnSerTYrMetAspTYrGlyPheTYrIleuValGlnMetGIYAspLeuSerGIY 123
 DB 345 GATTGGAAAGAAATATGTCCTAGCAGAACTTATATCATCTAGGATGATTCGCCCT 286

QY 124 AlaLeuIleTYrPheAspIYsProSerArgAlaIleGIYrGLuGIYArGValValAla 143
 DB 285 -----TTTAATCAACTATTACTGACGGGTAGAAACTGAGAAAGCCCAAGC 238

QY 144 IleGlnMetAlaTYrIleTYrHisGlnTYrGlnAlaAlaIYsSerProThrIYs 163
 DB 237 GATCAGAGATGGCTATAGTGTGAAGATACGAGAGAGACTCCAGCTTCAACAG 178

QY 164 AspAspTYrAsnAlaIYsSerAlaLeuGIYArGAlaIleSerGIYThr----- 181
 DB 177 GAACCT-----TCAGAGAACCCATTCCTGAAGATGACATGTTAGAAAACGGGTGT 124

QY 182 -----GlnHisapGIuIleIYsSer 189
 DB 123 AAGAGAAAGAGACCCCTCAAGATGCAAGAAAGGAAACAGAGCTGAGAGAGATT 64

QY 190 TYrAspIYsLeuLeuSerAspTYrIYsLeuLeuSerAspTYrIYs 204
 DB 63 ACATCCAAAGCTTATGACAAATACAGAGAGCTGCCATTATTA 19

RESULT 23
 LOCUS BQ299342 605 bp mRNA linear EST 16-MAY-2002
 DEFINITION sac46401.y1 Gm-cl054 glycine max cdna clone SOYBEAN CLONE ID:
 Gm-cl054-7082.5' similar to SW:OCT1 RAT P56558
 UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE
 110 KDA SUBUNIT ; mRNA sequence.

ACCESSION BQ299342
 VERSION BQ299342.1 GI:20814864
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

REFERENCE 1 (bases 1 to 605)
 Glycine.
 Shoemaker R., Keim P., Vodka L., Erpelting J., Coryell V., Khanna A., Holla B., Merrin M., Hillier L., Kucaba T., Martin J., Beck C., Wylie T., Underwood K., Stepien M., Theising B., Allen M., Bowers Y., Peterson B., Swaller T., Gibbons M., Pape D., Harrey N., Schurk R., Rittner E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann R., Waterston R. and Wilson R.
 Public Soybean EST Project
 Unpublished (1999)
 CONTACT: Shoemaker R./Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: cduresgen.com web site: www.resgen.com
 Seg primer: -40RP from Gibco
 High quality sequence stop: 421.

FEATURES
 source
 1..605
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl054-7082"
 /clone_1ib="Gm-cl054"
 /tissue_type="leaf, 3 week old, greenhouse grown"
 /lab_host="DH10B"
 /note="Vector: pBluescript II SK+, Site_1: EcoRI, Site_2: XhoI; The Harosoy NIL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht, 1995). The cDNA library was constructed from mRNA isolated from leaf tissue at various developmental stages of 3 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in cooperation with Dr. Paul Keim's laboratory at Northern Arizona University."

BASE COUNT 171 a 129 c 131 g 174 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.0194 Length: 605
 Score: 96.50 Matches: 39
 Percent Similarity: 41.71% Conservatives: 34
 Best Local Similarity: 22.29% Mismatches: 87
 Query Match: 9.25% Indels: 15
 DB: 14 Gaps: 4

US-10-069-544-2 (1-204) x BQ299342 (1-605)

QY 11 MetAlaMetGlyLeuValLeuSerAlaCysGlnSerThrProIleProProLysAsnAsn 30
 DB 78 ATGGGTAAGGAGACCTCACTGAGAGCCAGCTGTGCTGCAAGACTTGCATTAAT 117
 DB 31 ProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLysLeu 50
 DB 138 CCTCTATATGTTGATGCGCATAGCACTTGGGTATCTTATGAAGCTCAAGCTTGCTG 197
 QY 51 AspGlnAlaLysGlnGlnLeuAspAlaLeuSerAlaAspArgGlnIleAlaProAla 70
 DB 198 CAAGAAACATACAGTGTCTTACCTTGAGCATTTGGCATTCACCTTACCTTGGCTATGCA 257
 QY 71 TTAArgThrLeuAlaLysValIleTyrGlnAlaSerGlnAspAlaThrHisGlnThrLysAla 90

DB 258 TGGTCAATCTTCTGCTGCGCTTCATGAGACTCGGATTC-----AACAGAGCC 308
 QY 91 GlnArgLeuPheGluLysAlaIleGluLeuAsnProLysAspMetGlnSerTyrMetAsp 110
 DB 309 CTTCAGATATCAAG 368
 QY 111 TTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 129
 DB 369 CTTCGAGATGTATACAAAGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
 QY 130 -----LysProSerArgAlaIleGlyTyrGlnLysArgValAlaIleGlu 145
 DB 429 GCTTTCAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 485
 QY 146 -----AsnMetAlaTyrIleTyrTyrHisGlnTyrGlnAlaAlaLys 159
 DB 486 TATGAGCAAGGTCAACTGATATGCAATTCATCACTATTAACAAGCTGTTCATGTGAC 545
 QY 160 SerProThrLysAspAspTyrAsnAsnAlaLysSerAlaLeuGlu 174
 DB 546 CCCAGATTTTGGAGAGCTTACATTAACCTTGGGTATATCTCTAANA 590

RESULT 24
 BO870586 671 bp mRNA linear EST 15-AUG-2002
 BO870586
 LOCUS
 DEFINITION OGD9H11.yg ab1 OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
 OGD9H11, mRNA sequence.
 ACCESSION BO870586
 VERSION BO870586.1 GI:22257126
 KEYWORDS EST.
 SOURCE
 ORGANISM Lactuca sativa.
 Lactuca sativa.
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Astereaceae; Lactuceae; Lactuca.
 REFERENCE 1 (bases 1 to 671)
 Kozik A., Michelmore, R.W., Knapp S., Matvienko, M., Rieseberg, L., Lin, H., van Damme M., Lavelle, D., Chevalier, P., Ziegler, D., Blitson, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Composite Genome Project
 http://comgenomics.ucdavis.edu/
 Unpublished (2002)
 JOURNAL
 CONTACT: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Assumundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
 Singletron, see http://cgdb.ucdavis.edu/ for details.
 Plate: QGD9 row: H column: 11.
 FEATURES
 source
 1..671
 /organism="Lactuca sativa"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QGD9H11"
 /clone_1ib="OG_ABCDI lettuce salinas"
 /lab_host="E.coli"
 /note="Vector: pBluescript II, The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/TAG_LIB-OG_ABCDI lettuce salinas
 TAG_TISSUE=chemical induction

BASE COUNT 205 a 113 c 157 g 195 t 1 others
 ORIGIN TAG_SEQ=GTAGCCGCG"

Alignment Scores:
 Pred. No.: 0.0231 Length: 671
 Score: 96.50 Matches: 37
 Percent Similarity: 41.86% Conservatve: 17
 Best Local Similarity: 28.68% Mismatches: 58
 Query Match: 9.25% Indels: 17
 DB: 14 Gaps: 5

US-10-069-544-2 (1-204) x BQ870586 (1-671)

QY 30 AenProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
 DB 296 AATCCCATTTGCTGAGGATGATATATCTTGAGTTATATTAAGATAGACAAAT 355
 QY 50 LeuApglnAlaLysGlnGlnLeuAspAlaIleuSerAlaAspArgGlnPheAlaPro 69
 DB 356 CTGTATTAAGAGAGTGTATCATGTTGGCCCTATCATACAGCCCAATTTCTCCAG 415
 QY 70 AlaTyrArgThrLeuAlaLysValTyr-----GlnAlaSerGlnAspAlaThrHisGln 87
 DB 416 TCTTGAACACCTTGAGTACTTACACAGTTCAAGGTTAAATGATGCT----- 466
 QY 88 ThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAspProLysAspMetGlnSer 107
 DB 467 -----GCTGTCAGCATGTATGAAAAGCATTTGGCCAAATCCACGTATGCTGAGCT 520
 QY 108 TyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeu----- 125
 DB 521 TACATTAATCTTAGGGGTTCTTACAGAGATGCTGAGATATCTTTGGCCATTGAACA 580
 QY 126 -----IleTyrPheAspLysProSerArgAlaIleGlyTyrGlnLysArgVal 141
 DB 581 TATGAACAGTGCCTTAAATATAGACCCAGATTTCTGTAATGCCGCGC---CAGATTCAGATTA 637
 QY 142 ValAlaIleGlnAspMetAlaTyrIle 150
 DB 638 CTGTC-----ATGACATCAATT 655

RESULT 25
 D34850 327 bp mRNA linear EST 08-AUG-1994
 LOCUS CEK013C2P Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA
 DEFINITION clone yk13c2 5', mRNA sequence.
 ACCESSION D34850
 VERSION D34850.1 GI:526355
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae
 ; Rhabditidae; Peleiderinae; Caenorhabditis.
 REFERENCE Kohara,Y., Mitsuki,H., Nishigaki,A., Mochizuki,T., Sugimoto,A. and
 Tabara,H.
 TITLE Toward an expression map of the C.elegans genome
 JOURNAL Unpublished (1994)
 COMMENT Contact: Yui Kohara
 Genome Biology Lab.
 National Institute of Genetics
 Yata 111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.
 Location/Qualifiers
 1..327
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="yk13c2"
 /clone_lib="Yui Kohara unpublished cDNA"

BASE COUNT 96 a 57 c 74 g 99 t 1 others
 ORIGIN /sex="hermaphrodite, male"
 /tissue type="whole animal"
 /dev stage="varied"

Alignment Scores:
 Pred. No.: 0.00946 Length: 327
 Score: 95.50 Matches: 26
 Percent Similarity: 50.00% Conservatve: 17
 Best Local Similarity: 30.23% Mismatches: 34
 Query Match: 9.16% Indels: 9
 DB: 14 Gaps: 2

US-10-069-544-2 (1-204) x D34850 (1-327)

QY 30 AenProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyLys 49
 DB 5 AATCCTGATTTGTTGTTGTCAGAAATGATCTTGAAACTTACTTAAAGCAATGGAAGA 64
 QY 50 LeuApglnAlaLysGlnGlnLeuAspAlaIleuSerAlaAspArgGlnPheAlaPro 69
 DB 65 CTGTGAAGAGCGAGGTTGTTACTTGAAACCAATCGAACTCAACACAGATCGCTGTC 124
 QY 70 AlaTyrArgThrLeuAlaLysValTyrGlnAlaSerGlnAsp-----AlaThrHis 86
 DB 125 GCATGTCATCTTGCATGTGTATTCATATGTCAGAGAAATTTGGTGGCATTCAT 184
 QY 87 GlnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlnLeuAspProLysAspMetGln 106
 DB 185 CAT-----TTGAGAAAGCTGTACTTGTGATTCAAACTTCCTCGAC 226
 QY 107 SerTyrMetAspTyrGly 112
 DB 227 GCTTATATTAATCTTGA 244

RESULT 26
 CNS07DGB 1014 bp DNA linear GSS 08-JUL-2001
 LOCUS T3 end of clone BD0A012D06 of library BD0A from strain CBS 94 of
 DEFINITION Candida tropicalis, genomic survey sequence.
 ACCESSION AL440433
 VERSION AL440433.1 GI:12223844
 KEYWORDS GSS.
 SOURCE Candida tropicalis.
 ORGANISM Candida tropicalis.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 REFERENCE 1 (bases 1 to 1014)
 AUTHORS Boicet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 de-Montigny,J., Dujon,B., Durren,P., Lepingle,A., Liorette,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
 MEDLINE 20584711
 PUBMED 1152876
 2 (bases 1 to 1014)
 Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
 Dujon,B.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 16. Candida
 tropicalis
 JOURNAL FEBS Lett. 487 (1), 91-94 (2000)
 MEDLINE 20584726
 PUBMED 1152891
 3 (bases 1 to 1014)
 Genoscope.
 TITLE Direct Submission
 REFERENCE Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :


```

Db      331  AGCAATCTGGCTTCATTCACAAAGATTCCAGGAATATCCAGAAACATACCTTCTTAC 390
Qy      153  HHSGLNTYGLIALALALysSerProthrlYsAPAPPTyAsnAsnAlALysSerAla 172
Db      391  CCCACAGCTCTGAAACTTAAAGCTTGACTTCTCTGATGTTATGTTACTTGCTCATTTGC 450
Qy      173  LeuGLIARGLALALeulIeSerGLYThrGlnHISAPGLIleLysSerTyRAspLys 192
Db      451  CTACAG---ATTGCTGTGATGATGACAGACTATGATGACGATGAG-----AAA 498
Qy      193  LeuLeuSer 195
Db      499  TTGGTTAGT 507

RESULT 28
CNS0756E      917 bp DNA linear GSS 07-JUL-2001
LOCUS      clone BA0AB033C05 of library BA0AB from strain CLIB 210 of
DEFINITION  Kluveromyces lactis, genomic survey sequence.
ACCESSION  AL429708
VERSION    AL429708.1 GI:12212902
KEYWORDS   GSS.
SOURCE     Kluveromyces lactis.
ORGANISM  Kluveromyces lactis.
REFERENCE  1 (bases 1 to 917)
AUTHORS   Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
          Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
          de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
          Malperuy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
          Saudin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
          Wincker,P. and Weissenbach,J.
          Genomic exploration of the hemiascomycetous yeasts: 1. A set of
          yeast species for molecular evolution studies
          FEBS Lett. 487 (1), 3-12 (2000)

TITLE
JOURNAL    2 (bases 1 to 917)
MEDLINE    Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F.,
PUBMED     Duchateau-Nguyen,G., Lemaire,M., Marnelisse,R., Montrochet,R.,
20584711    Robert,C., Termlier,M., Wincker,P. and Wesolowski-Louvel,M.
REFERENCE   Genomic exploration of the hemiascomycetous yeasts: 11.
AUTHORS     Kluveromyces lactis
TITLE       FEBS Lett. 487 (1), 66-70 (2000)
JOURNAL     20584721
MEDLINE     11152886
PUBMED      3 (bases 1 to 917)
REFERENCE   11152886
AUTHORS     Genoscope.
TITLE       Direct Submision
JOURNAL     Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
          2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail:
          beq@genoscope.cns.fr - Web: www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
          yeast species: Saccharomyces byanus var. uvarum, Saccharomyces
          exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
          Saccharomyces kluyveri, Kluveromyces thermotolerans, Kluveromyces
          lactis var. lactis, Kluveromyces marxianus var. marxianus, Pichia
          angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
          Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
          5 kb were prepared and both extremities were sequenced. See
          keywords for description of this sequence and for the sequence of
          the other extremity of this insert.
          Location/Qualifiers
            1..917
              /organism="Kluveromyces lactis"
              /strain="CLIB 210"
              /variety="lactis"
              /db_xref="taxon:26985"
              /clone="BA0AB033C05"
              /clone_11b="BA0AB"
              misc_feature
                <2..>715

```

```

BASE COUNT  274 a 175 c 210 g 258 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0628          Length: 917
Score: 95.00             Matches: 40
Percent Similarity: 38.31% Conservative: 19
Best Local Similarity: 25.37% Mismatches: 53
Query Match: 9.11%        Indels: 42
Db: 17                    Gaps: 5

US-10-069-544-2 (1-204) x CNS0756E (1-917)

Qy      49  LysLeuAPRGLIALALysGlnGlnLeuAPRAlALeUSerAlaSP----- 64
Db      47  AAATTGTGAGGCTATTGAAATAATACACTCTCTATCGAAGTTTCTCATTAAGCCT 106
Qy      65  -----ArgGlnPheAlaProAlaTyRArgThrLeuAlALysValTyRGLNALSer 81
Db      107  GTTACTACTGGAACAGGCTCGACGCTTATCTTCTTGAAGCAGTACGAACAGGCCCTT 166
Qy      82  GlnAPRAlaThrHISGlnThrLysAlaGlnArgLeuPheGlnTyRAlaIleGlnLeuAsn 101
Db      167  AAGAGCGC-----GAGCAAGCATATGAAGTTAT 196
Qy      102  ProLysAPRMetGlnSerTyRMetAspTyRGLYpHeTyRLeuValGlnMetGLYAspLeu 121
Db      197  CCTCATATTTCTAAGGGGTTTTCTAGACTTGATTT----- 232
Qy      122  SerGlnAlaLeuIleTyRPhAPRProSerArgAlaIleGLYTyRGLUGLYArgVal 141
Db      233  -----GCGAAGTATGCTTTGAAATAGGCTTAAGAAAGCTTTAGATGCTTACAGAAAGCTC 286
Qy      142  ValAlaIleGlnAsnMetAlaTyRILETyRThrHISGlnTyRGLUALALALysSerPro 161
Db      287  CTAGATTCGAAGG-----GAAAAGCTACGAGCTT 319
Qy      162  ThrLysAPRAspTyRAsnAsnAlALysSerAla-----LeuGLIARg 175
Db      320  ATGAAAGAGATTGCGAAATCGCAAGAGATAGAACTCATGATGATGATGAGAG 379
Qy      176  ALALeulIeSerGLYThrGlnHISAPGLIleLysLysSer 189
Db      380  AGCTCTACAGCGGCTTCTTGAAGATGTTTCTGACAAATCT 421

RESULT 29
BO434018      919 bp mRNA linear EST 24-MAY-2002
LOCUS      AGENCOURT 7760927 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5062020
DEFINITION  5', mRNA sequence.
ACCESSION  BO434018
VERSION    BO434018.1 GI:21173094
KEYWORDS   EST.
SOURCE     human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 919)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
          unpublished (1999)
JOURNAL
COMMENT    Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: ATCC/DCPD/DPF
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LMU at:
          http://image.llnl.gov

```

/note="similar to Saccharomyces cerevisiae ORF YOR007c [SGT; similarity to protein phosphatases]"

/evidence=not experimental

FEATURES	source	misc_feature	BASE COUNT	ORIGIN
AUTHORS	Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Artiguenave,F., Gallardin,C. and Casaregola,S		314 a 143 c 137 g 390 t	3 others
TITLE	Genomic exploration of the hemiascomycetous yeasts: 6.			
JOURNAL	Saccharomyces exiguus			
MEDLINE	FEBS Lett. 487 (1), 42-46 (2000)			
PUBMED	20584716			
REFERENCE	11152881			
AUTHORS	3 (bases 1 to 987)			
TITLE	Genoscope.			
JOURNAL	Direct Submission			
COMMENT	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angustis, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert. Location/Qualifiers			
FEATURES	1..987			
source	/organism="Saccharomyces exiguus" /strain="CBS 379" /db_xref="taxon:34358" /clone="AY00A010H08" /clone_11b="AY00A" /note="end : 13" complement(<3..>986) /note="similar to Saccharomyces cerevisiae ORF YHR117w [TOM71 : strong similarity to Tom70p/Mas70p] similar to Saccharomyces cerevisiae ORF YNL121c [TOM70 ; mitochondrial outer membrane specialized import receptor]" /evidence=not_experimental			
BASE COUNT	314 a 143 c 137 g 390 t			3 others
ORIGIN				
Alignment Scores:				
Pred. No.:	0.0832	Length:	987	
Score:	94.50	Matches:	45	
Percent Similarity:	35.12%	Conservative:	27	
Best local Similarity:	21.95%	Mismatches:	60	
Query Match:	9.06%	Indels:	73	
DB:	17	Gaps:	9	
US-10-069-544-2 (1-204) x CNS06P2E (1-987)				
Oy	27	ProlybAaAaAaAaPfcGlnLeuAlaGln-----	35	
Db	911	CGAGAAAGATTAACCAATGATGTTCCACAAATTTMCCTCAATATCGCAATGCGTTCAATTC	852	
Oy	36	-----lIaevThrcGlnleAlaIleSeerLeuLeuAepMecGly-----	48	
Db	851	TTCCGTATCTTTAAACCTGAAGTAACCTCGATTAACCTTGACGAAATTAATGAAAAACGAT	792	
Oy	49	---LysLeuAepGlnAlaLysGlnGlnLeuAaAlaAla-----Leuser 62		
Db	791	AAATCTTTACAAAGCGTTTACAGATTTATTTCTGCTACTTAAGAACGTTATTTATTA	732	
Oy	63	AlaAaPaRgGlnPheAlaProAlaTyraRgThrLeuAlaLysValTyrglnAlaSerGlu 82		
Db	721	GCTGAAAAATTAATTTCAAAATGCTGAAGAAGGTTATTTAAACAAATTTAATGATACTAAT	672	
Oy	83	AepAlaThrHicGlnThrLys-----	89	
Db	611	CAATGAAATCTATTAAACAAATTAACCAATGCTTTGAATTAATTTCTGGATCTTTAAATTT	612	
Oy	90	-----AlaGlnAArgLeuPheGluValAlaIleGlnLeuAaAaAaPfcGlnLeuAlaGln-----	102	

Db	611	TTACTATGATTGTTAACTGGGTCCTCAGAAAGATTAAAAAGCATTTGATTATTCCT	552
Qy	103	LyasapmcclnserTyrmecAspTyrglyPheTyrlenuValGlnMetGlyAspleuSer	122
Db	551	CGT---GTTAAATTCCTTATCTATATATGAGCATTTAAACAATGCGCTGATTAAGAGTAACTCA	495
Qy	123	GlyAlaIeuLeuIeTyrlPheAspIyProSerArgAlaIleGlyTyrgluGlyArgValVal	142
Db	494	GATTATTTCACTATTTGATAAA-----	471
Qy	143	AlaIleGlu-----AsnMetAlaTyrlIeTyrlrHis-----	153
Db	470	GCTTAAAGATTAGATCCAAATGTCTTCGCTACTATATTCATTAAGGTCATTAATTC	411
Qy	154	-----GlnTyrgluAlaAlaLysSerProThrLysAspAspTyrlAsnAsnAlaLys	170
Db	410	ATTACTCAAAAATTCGATTAGCAAGAAAG-----GATTTCATTTAAAGCTTAA	363
Qy	171	SerialenGluArg	175
Db	362	GAACCTGATGAACA	348
RESULT 34			
LOCUS	BO070967	1001 bp	mRNA linear EST 02-APR-2002
DEFINITION	AGENCOURT_6855903 NIH_MGC_47 Homo sapiens CDNA clone IMAGE:5923457		
ACCESSION	BO070967		
VERSION	BO070967.1	GI:19900013	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 1001)		
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-r@mail.nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Rubin Laboratory		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: LCM2091 row: p column: 18		
FEATURES	High quality sequence, stop: 726.		
source	Location/Qualifiers		
	1..1001		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5923457"		
	/clone_id="NIH_MGC_47"		
	/tissue_type="neuroblastoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: brain; Vector: pOT81; Site 1: XhoI; Site 2:		
	EcoRI; cDNA made by oligo-dT priming. Directionally		
	cloned into EcoRI/XhoI sites using the following 5'		
	adaptor: GGCACGAG(G). Size-selected >500bp for average		
	insert size 1.8kb. Library constructed by Ling Hong in		
	the laboratory of Gerald M. Rubin (University of		
	California, Berkeley) using ZAP-cDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies).		
	Note: This is a NIH_MGC Library."		
BASE COUNT	198 a 316 c 302 g 184 t	1	others
ORIGIN			
Alignment Scores:			
Pred. No.:	0.0852	Length:	1001
Score:	94.50	Matches:	52
Percent Similarity:	40.38%	Conservative:	32

Best Local Similarity: 25.00%
 Query Match: 9.06%
 DB: 14
 Gaps: 12

US-10-069-544-2 (1-204) x B0070967 (1-1001)

```

QY 47 MetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaAlaLeuSerAlaAspArgGln 66
DB 198 ATGAACGAGCTCGGCTCGAGCCAGCTGTGTGACGTACACCTGACAGTCAAGTACAG 257
QY 67 PheAlaProAlaTyrArgThrLeuAla---LysValTyrGlnAlaSerGlnAspAla--- 84
DB 258 GATGACCGCGCCCGGAGTTCATGCGCCCAAGGTCGTGCTGCTCATCCAGCCCTGTC 317
QY 85 -----ThrIsglnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlu 99
DB 318 TTCAGGCCCATGTTACCAACGAGGCTGCGGAGCAGAGCATGAGGTGTGTCCATTGAG 377
QY 100 ---LeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeu---ValGln 117
DB 378 GGTATCCACCCCAAGGTCTATGAGCGCTCATTAATTGCGCTACACGCGCTCCATCTCC 437
QY 118 MetGlyAsp-----LeuSerGlyAlaLeuIleTyr---PheAspLys 130
DB 438 ATGGCGGAGAAAGTGTCTCTCCACGTACAGACGCTGTCTCATGATACAGATTCAGAC 497
QY 131 -----ProSerArgAlaIleGly 136
DB 498 GTTGTCCGTGCTGAGTACTTCTGTGACAGACAGTGGACCCCAATGCCATCGCGC 557
QY 137 TyrGluGlyArgValValAlaIleGlnAspMetAla----- 148
DB 558 -----ATCGCCAACTTCTGCTGAGCAGACAGTTGCTGTGAGAGTTG 596
QY 149 -----TyrIleTyrThrIsglnTyrGlnAlaAlaLysSerPro--- 161
DB 597 CACCAAGCGTCCCGGAGTATCATCTACATGCAATTTTGGGAGGTGGCCAAAGAGAG 656
QY 162 -----ThrLysAspAspTyrAsn--- 167
DB 657 TTCTTCAACCTGTCCCACTGCACTGACCTGATGACCGCGGAGCAGCTGAACGTG 716
QY 168 AsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGlyThrGlnHisAsp---GluIle 186
DB 717 CGCTGCGAGTCCGAGCTTCCACGCTGCATCACTGGGTCAAGTACGATCGCAACAG 776
QY 187 LysLysSerTyrAspLysLeuLeu 194
DB 777 CGACGGTTCACGTCACGCGCTG 800

```

RESULT 35
 B0070453 1015 bp mRNA linear EST 02-APR-2002
 LOCUS AGENCOURT 6839432 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5922839
 DEFINITION 5', mRNA sequence.
 ACCESSION B0070453
 VERSION B0070453.1 GI:19899499
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1015)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: gsaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LNCM2090 row: f column: 24
 High quality sequence stop: 655.
 Location/Qualifiers
 1. 1015

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5922839"
 /clone_lib="NIH_MGC_47"
 /tissue_type="neuroblastoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 197 a 322 c 305 g 191 t
 ORIGIN

Alignment Scores:

Pred. No.: 0.0872 Length: 1015
 Score: 94.50 Matches: 52
 Percent Similarity: 40.38% Conservative: 32
 Best Local Similarity: 25.00% Mismatches: 57
 Query Match: 9.06% Indels: 67
 DB: 14 Gaps: 12

US-10-069-544-2 (1-204) x B0070453 (1-1015)

```

QY 47 MetGlyLysLeuAspGlnAlaLysGlnLeuAspAlaAlaLeuSerAlaAspArgGln 66
DB 198 ATGAACGAGCTCGGCTCGAGCCAGCTGTGTGACGTACACCTGACAGTCAAGTACAG 257
QY 67 PheAlaProAlaTyrArgThrLeuAla---LysValTyrGlnAlaSerGlnAspAla--- 84
DB 258 GATGACCGCGCCCGGAGTTCATGCGCCCAAGGTCGTGCTGCTCATCCAGCCCTGTC 317
QY 85 -----ThrIsglnThrLysAlaGlnArgLeuPheGlnLysAlaIleGlu 99
DB 318 TTCAGGCCCATGTTACCAACGAGGCTGCGGAGCAGAGCATGAGGTGTGTCCATTGAG 377
QY 100 ---LeuAsnProLysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeu---ValGln 117
DB 378 GGTATCCACCCCAAGGTCTATGAGCGCTCATTAATTGCGCTACACGCGCTCCATCTCC 437
QY 118 MetGlyAsp-----LeuSerGlyAlaLeuIleTyr---PheAspLys 130
DB 438 ATGGCGGAGAAAGTGTCTCTCCACGTACAGACGCTGTCTCATGATACAGATTCAGAC 497
QY 149 -----TyrIleTyrThrIsglnTyrGlnAlaAlaLysSerPro--- 161
DB 597 CACCAAGCGTCCCGGAGTATCATCTACATGCAATTTTGGGAGGTGGCCAAAGAGAG 656
QY 162 -----ThrLysAspAspTyrAsn--- 167
DB 657 TTCTTCAACCTGTCCCACTGCACTGACCTGATGACCGCGGAGCAGCTGAACGTG 716
QY 168 AsnAlaLysSerAlaLeuGlnArgAlaLeuIleSerGlyThrGlnHisAsp---GluIle 186
DB 717 CGCTGCGAGTCCGAGGTCTTCCACGCTGCATCACTGGGTCAAGTACGATCGCAACAG 776

```

Qy 187 LysylSerTyrAspLysLeuLeu 194
 DB 777 CGACGGTCTACGTCACGCGCTG 800

RESULT 36
 CDS06M7H 1051 bp DNA linear GSS 17-JUN-2001
 LOCUS T3 end of clone AU0AA002F01 of library AU0AA from strain CBS 3082
 DEFINITION of Saccharomyces kluyveri, genomic survey sequence.

ACCESSION AL405123
 VERSION AL405123.1 GI:12167477
 KEYWORDS GSS.

SOURCE Saccharomyces kluyveri.
 ORGANISM Saccharomyces kluyveri.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 1051)
 AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 deMontigny,J., Dujon,B., Durieux,P., Lepingle,A., Llorente,B.,
 Mauryeruy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Porter,S.,
 Sautin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissesbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies

TITLE PBBS Lett. 487 (1), 3-12 (2000)

JOURNAL MEDLINE 20584711
 PUBMED 11152876
 AUTHORS 2 (bases 1 to 1051)
 TITLE Neugeglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
 Gallardin,C. and Casaregola,S.
 Genomic exploration of the hemiascomycetous yeasts: 9.
 Saccharomyces kluyveri
 PBBS Lett. 487 (1), 56-60 (2000)

JOURNAL MEDLINE 20584719
 PUBMED 11152884
 AUTHORS 3 (bases 1 to 1051)
 TITLE Genoscope.
 Genoscope.
 Direct Submission
 Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 beqet@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermocolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES
 source
 1..1051
 /organism="Saccharomyces kluyveri"
 /strain="CBS 3082"
 /db_xref="taxon:4934"
 /clone="AU0AA002F01"
 /clone_1fb="AU0AA"
 /note="end : T3"
 /note="and : T3"
 misc_feature
 <5..>636
 /note="similar to Saccharomyces cerevisiae ORF YOR007C [
 SGT2, similarity to protein phosphatases]
 1 putative frameshift(s)
 /evidence="not_experimental"

BASE COUNT 286 a 217 c 314 t 2 others

ORIGIN

Alignment Scores:
 Pred. No.: 0.0925 Length: 1051
 Score: 94.50 Matches: 39
 Percent Similarity: 40.58% Conservative: 17

Best Local Similarity: 28.26% Mismatches: 47
 Query Match: 9.06% Indels: 35
 DB: 17 Caps: 5

US-10-069-544-2 (1-204) x CDS06M7H (1-1051)

Qy 58 AspaAaLeuSerAlaAspaRglnPheAlaProAlaTyrArgThrLeuAlaVal 77
 Db 62 AATGCTGTTATTACGCCACAGG-----GCTGCCGCTATTCTCTTGAAGAAATAC 115
 Qy 78 TyrGlnAlaSerGluAspaAlaThrHisGlnThrIysAlaGlnArgLeuPheGluVal 97
 Db 116 GAGGAAATCCGTTCAAGATGCT-----GAGTCTGCT 145

Qy 98 IlleGluLeuAnpPolysAspMetGlnSerTyrMetAspTyrGlyPheTyrLeuValGln 117
 Db 146 ATTAGGGAACCCCTTCTTATTCAGAGGGTACTTGAATTGATTAAGATTGCAAGATAGCT 205

Qy 118 MetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysProSerArgAlaIleGlyTyr 137
 Db 206 CTAGGT-----AAGCCAGAAAGCTTTGAAGACC 235

Qy 138 GluGlyArgValValAlaIleGluAsnMetAlaTyrIleTyrTyrHisGlnTyrGlnAla 157
 Db 236 TACAAACGCTTTTGACATTCAGAGGC-----GACCAA 268

Qy 158 AlaLysSerProThrIysAspaSptTyrAsnAlaLysSerAlaLeuGluArgAlaLeu 177
 Db 269 GCCACTGATGTCATGAAGAGATTCAGAACTGCCAGAGAAAGTTGAACAATCTTTG 328

Qy 178 -----IleSerGlyThrGlnHisAspGluIleTyrLysSerTyrAsp 191
 Db 329 AACTTGAAAGAGATGCCAACACAAACGATGACAGAGACGACACCGAT 382

RESULT 37
 BM952706 468 bp mRNA linear EST 14-MAR-2002
 LOCUS BM952706
 DEFINITION 952057D09.xl 952 - BMS tissue from Walbot Lab (reduced rRNA) Zea
 mays cDNA, mRNA sequence.

ACCESSION BM952706
 VERSION BM952706.1 GI:19436296
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 468)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 JOURNAL Unpublished (1999)
 COMMENT
 CONTACT: Walbot V.
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 952057 row: D column: 09.

FEATURES
 source
 1..468
 /organism="Zea mays"
 /cultivar="BMS (Black Mexican Sweet)"
 /db_xref="taxon:4577"
 /clone_1fb="952 - BMS tissue from Walbot Lab (reduced rRNA)"
 /tissue_type="suspension culture"
 /dev_stage="mixed logarithmic and stationary growth
 phases"
 /lab_host="DH10B"
 /note="Vector: pUC19, Site_1: EcoRI, Site_2: EcoRI, The
 library was prepared by George Rudenko using poly (A)

Qy	154	-----GmTYRGJLALALALysSerProrHrlyvSaPaPpTYrAsnaAlaAlaYsSer	171
Db	447	CCTCGGAATATGAAAGCTGCACAAAGAGAGGATTAATAAAAAAGAGCGTACAGG	506
Qy	172	AlaLeuGuaRgaAlaLeuIeSerGlyThrGlnHlaePguIuIeLySeTYrAaP	191
Db	507	CAGCGCCCGCTGT-----GAAAGCTCAGGCTGGCTATAG	542
Qy	192	lys	192
Db	543	AAG	545
RESULT 39			
LOCUS	BM11682	490 bp	mRNA linear
DEFINITION	EST559218 potato roots Solanum tuberosum cDNA clone CPRO1215 5' end		
ACCESSION	BM11682		
VERSION	BM11682.1	GI:17074412	
KEYWORDS	EST.		
SOURCE	Potato.		
ORGANISM	Solanum tuberosum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; easteride I; Solanales; Solanaceae; Solanum.		
AUTHORS	1 (baes 1 to 480) van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S., Uteerbek,T., Chelmingo,A., Bougri,O., Buehl,C.R., Romning,C., Tanksey,S. and Baker,B. Generation of ESTs from potato roots Unpublished (2001)		
TITLE	CONTACT: Research Genetics, Libraries Division		
JOURNAL	CONTACT: Research Genetics, Libraries Division		
COMMENT	TELE: 1-800-711-6195 Email: cdna@resgen.com For clone info: please contact Research Genetics, Libraries Division tel: 1-800-711-6195, email cdna@resgen.com Seq primer: T3		
FEATURES	Location/Qualifiers		
source	1..490 /organism="Solanum tuberosum" /cultivar="Kennebec" /db_xref="taxon:4113" /clone="CPRO1215" /clone_1lb="potato roots" /tissue_type="roots" /dev_stage="in vitro grown stem cuttings" /lab_host="SOLR" /note="vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Coriell University, Tanksey Lab; sequencing: The Institute for Genomic Research. Roots were isolated from in vitro grown stem cuttings on CM medium. Roots were isolated two weeks after placing the stem cuttings from in vitro grown plants on medium."		
BASE COUNT	153 a	87 c	108 g 142 t
ORIGIN			
Alignment Scores:			
Pred. No.:	0.0351	Length:	490
Score:	93.50	Matches:	29
Percent Similarity:	44.66%	Conservative:	17
Best Local Similarity:	28.16%	Mismatches:	50
Query Match:	8.96%	Indels:	7
DB:	13	Gaps:	2
US-10-069-544-2 (1-204) x BM11682 (1-490)			
Qy	30	AenProGlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuAaPMeClyLys	49
Db	168	AATCCTCATTTGGCTGGAAGCATGCAACATTGTCGAGCATATACAAAGATAGGACAAC	227
Qy	50	LeuAaPGuIaLyGlnLeuAaPAlaAlaLeuSerAlaAaPArgGlnPheAlaPro	69
Db	228	CTTGATTAAGCGCTGAATGTTATTCAGATGGCTTTGTCAATCAACCAAGCTTCTCAG	287

QY	70	AlaTyrArgThrLeuAlaValTyr-----GlnAlaSerGluAspAlaThrIleGln	87
DB	288	TCATTAAACATCTTCGGAGTAGTTTACACTGTTACGGGTAATAATGATGCT-----	338
QY	88	ThrTyrAlaGlnArgLeuPheGlnIleValaIleGluLeuAsnProLysAspMetGlnSer	107
DB	339	-----GCTGTACGATCATTTGAGAAAGCTATTCATTGCAAAACCCAAATATGCAGAGGCA	392
QY	108	TyrMetAspTyrGlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaIleuTyr	127
DB	393	TATATATATTAGAGGTTCTTTACAGGATCAGGCAATATATCTCGGCATTGAAGA	452
QY	128	PheAspLys	130
DB	453	TATGAGCAA	461
RESULT 40			
LOCUS	AM622449	587 bp	mRNA linear EST 18-MAY-2000
DEFINITION	EST313237 tomato root during/after fruit set, Cornell University		
ACCESSION	LYCopersicon esculentum cDNA clone CLEX15A14 5', mRNA sequence.		
VERSION	AM622449		
KEYWORDS	AM622449.1 GI:7334084		
SOURCE	EST.		
ORGANISM	LYCopersicon esculentum		
REFERENCE	van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Kochian,L.,		
AUTHORS	Liang,F., Hansen,T., Craven,M.B., Bowman,C.L., Roming,C.M.,		
	Mierman,W., Frazer,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley		
	S.D.		
	Generation of ESTs from tomato root, during and after fruit set		
TITLE	Unpublished (1999)		
JOURNAL	Contact: CUGI		
COMMENT	Clemson University Genomics Institute		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Email: http://www.genome.clemson.edu/orders/index.html		
	5 prime sequence.		
FEATURES	Location/Qualifiers		
SOURCE	1..587		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA496"		
	/db_xref="taxon:4081"		
	/clone="CLEX15A14"		
	/clone_lib="tomato root during/after fruit set, Cornell		
	University"		
	/cissue_type="root"		
	/dev_stage="plants during and after fruit-set"		
	/note="Vector: pBluescript SK(-); Site 1: EcoRI, site 2:		
	XhoI; supplier: Tankeley; Tissue supplied by Dave Garvin		
	(USDA-ARS, Ithaca, NY 14850)."		
BASE COUNT	158 a 123 c 146 g 160 t		
ORIGIN			
Alignment Scores:			
Pred. NO.:	0.0476	Length:	587
Score:	93.50	Matches:	40
Percent Similarity:	40.61%	Conservative:	27
Best Local Similarity:	24.24%	Mismatches:	71
Query Match:	8.96%	Indels:	27
DB:	10	Gaps:	5
US-10-069-544-2 (1-204) x AM622449 (1-587)			
QY	12	AlaMetGlyLeuValLeuSerSerAlaCysGlnSerThrProIleProGlyAsnAsnPro	31
DB	6	GCACAAAGGTTTGGTGCAAGAGCATACATTTGATGTGAGAGCATCGCATACAACTT	65

```
QY 32 GlnLeuAlaGlnIleArgThrGlnIleAlaIleSerLeuLeuAspMetGlyValLeuAsp 51
Db 66 ACATTGGCTGTGACATGATGCTTGGCTTTCATGATGCTGGCATTTAC 125
QY 52 GlnAlaValysGlnGlnLeuAspAlaIleLeuSerAlaAspArgGlnPheAlaProAlaTyr 71
Db 126 AGGGCATTTGCATATTACAGAGAGCTGCAAGCTAAACCAATTTTCAGATGCATAT 185
QY 72 ArgThrLeuAlaValysValTyrGlnAlaSerGluAspAlaThrHisGlnThrLysAlaGln 91
Db 186 TTGAACCTGGAAATGTGTATAAGCT-----CTGGGATGCCACAGAGGCCATC 236
QY 92 ArgLeuPheGlnLysAlaIleGlnLeuAsnProLysAspMetGlnSerTyrMetAspTyr 111
Db 237 ATGTGTTACCAAGGCTCTCTCTGTCGACACAGAC----- 272
QY 112 GlyPheTyrLeuValGlnMetGlyAspLeuSerGlyAlaLeuIleTyrPheAspLysPro 131
Db 273 -----TATGCTGTGCTTTTGGCAACTTA-----GCTACTGTGTAATAATA----- 314
QY 132 SerArgAlaIleGlyTyrGlnGlyArgValValAlaIleGluAsnMetAlaTyrIleTyr 151
Db 315 -----CAAGGTAACTG-----GAGATGGCCATGCTTAAT 344
QY 152 TyrHisGlnTyrGlnAlaAlaValysSerProThrLysAspAspTyrAsnAsnAlaLysSer 171
Db 345 TACAGGAGAGCTTAACTGTGATGCTGATTTCTTGAGGCAATATAACAATCTGGGTAAAT 404
QY 172 AlaLeuGluArgAla 176
Db 405 GCTCTGAAAGATGCT 419
```

Search completed: July 4, 2003, 04:03:56
Job time : 1098 secs